



## **Essays on Networks and Labor Market Mobility**

by Ian Michael Schmutte

---

This thesis/dissertation document has been electronically approved by the following individuals:

Abowd, John Maron (Chairperson)

Blume, Lawrence Edward (Minor Member)

Kleinberg, Jon M (Minor Member)

Freedman, Matthew (Minor Member)

# ESSAYS ON NETWORKS AND LABOR MARKET MOBILITY

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Ian Michael Schmutte

August 2010

© 2010 Ian Michael Schmutte  
ALL RIGHTS RESERVED

## ESSAYS ON NETWORKS AND LABOR MARKET MOBILITY

Ian Michael Schmutte, Ph.D.

Cornell University 2010

This dissertation applies the tools of network analysis to study job mobility. Job mobility is a complex phenomenon, and network theory provides a novel and practical framework for dealing with this complexity in understanding how individuals move from job to job. My first essay measures the effect of job referral networks on search outcomes. The key contribution of this essay is providing evidence of one mechanism by which social interactions affect earnings. An on-the-job search model extended to include social transmission of job information yields an empirical specification in which ones current job offer depends on the average offer of his social contacts. Using block level variation in the quality of jobs held by ones residential neighbors, I find that when changing jobs an individual with better local network contacts will obtain a higher quality job. In addition to the main result, this paper provides new evidence on the spatial structure of the wage distribution within urban areas. In the second essay I apply network algorithms to detect groups of workers and employers with relatively homogeneous patterns of job mobility. Workers with interchangeable skills should have similar patterns of mobility across employers that use those skills in roughly the same way. Grouping workers and jobs solely on the basis of similar mobility patterns reveals labor market sectors with distinct compensation structures. My final essay, joint with John Abowd, uses network models to facilitate identification of employer-specific wage premia in a decomposition of log earnings from matched employer-employee data.

## **BIOGRAPHICAL SKETCH**

Ian Schmutte received his Bachelor's degree from Indiana University with a double major in Mathematics and Philosophy. Upon deciding to pursue a research career in economics, he enrolled in and completed the Master of Arts degree in Economics at the University of Missouri, St. Louis. From there, he spent a year doing research under a Fulbright fellowship in Australia on the response of labor unions in that country to increasing international trade. Through his studies, Ian began to articulate a longstanding interest in the connections between human behavior, social institutions, and the organization of labor in modern economies. His pursuit of these ideas led him to undertake and ultimately complete the doctoral degree in Economics at Cornell University.

For Desmond and Amelia.

## ACKNOWLEDGEMENTS

The research and writing of this dissertation required the intellectual, material, and moral support of an uncountable number of people and an endless number of events. I am first and foremost indebted to my dissertation advisor, John Abowd. John has been a steadfast mentor whose support and enthusiasm for my research has been unflagging. Every graduate student hopes for an advisor who will be part mentor, part colleague, and part friend. John has been all three, and I count myself among the lucky few who actually find this experience.

I owe an equal debt to the members of my dissertation committee, Larry Blume, Jon Kleinberg and Matthew Freedman. Larry always pushed for a deeper analysis and for the work to have a more fundamental contribution to economics. In addition, he provided crucial insights when I found myself stuck thinking about certain modeling problems. Jon taught me about the analysis of complex networks, and pointed out areas where tools from computer science could be brought to bear on the economic issues in my dissertation. Finally, Matt Freedman has been a continual source of advice on how to think about the complicated issues of job search in urban labor markets, and was always quick to return very insightful comments on my written work.

My parents and grandparents have provided unflagging emotional and moral support on what turned out to be a long journey. Being able to spend the time and to find the focus to do original research is a wonderful experience, but one that can be lonely and isolating. Somehow, my wife Tiffanie gave me the space to do research without the isolation. Because of her, and because our two children, Desmond and Amelia, were born along the way, I can truly say that graduate school was one of the greatest times of my life.

Many of my colleagues at Cornell and elsewhere provided feedback, support and guidance at various points along the way. I am especially grateful in this regard to my collaborators, Francis Kramarz and Kaj Gittings, and my many friends, colleagues and mentors including Jessica Bean, Chris Cotton, Brian Dillon, Jayant Ganguli, Amanda Griffith, Hope Michaelson, Emily Owens, Joseph Price, Mike Strain, Joshua Teitelbaum, and Russell Toth.

The research in this dissertation involved unique data and computing requirements, and in these regards, I have an enormous debt to the Longitudinal Employer Household Dynamics Program and the Center for Economic Studies at the U.S. Census Bureau. They providing material support for my doctoral studies as well as access to the data I needed to address the central questions of my dissertation. As importantly, I spent my doctoral studies in the company of a group of great research economists at Census. I owe a special thanks to Lars Vilhuber and Kevin McKinney who taught me most of what I know about the nuts and bolts of working with LEHD data, and more importantly, in thinking about them like an economist. I must also thank my many other colleagues at Census who supported my research, especially Lynn Riggs, Matt Graham, Lucia Foster, and Ron Jarmin.

This research uses data from the Census Bureau's Longitudinal Employer Household Dynamics Program, which was partially supported by the National Science Foundation Grants SES-9978093, SES-0339191 and ITR-0427889; National Institute on Aging Grant AG018854; and grants from the Alfred P. Sloan Foundation. Any opinions and conclusions expressed herein are those of the author and do not necessarily represent the views of the U.S. Census Bureau, its program sponsors or data providers, or of Cornell University. All results have been reviewed to ensure that no confidential information is disclosed.



## TABLE OF CONTENTS

Biographical Sketch . . . . .	iii
Dedication . . . . .	iv
Acknowledgements . . . . .	v
Table of Contents . . . . .	vii
<b>List of Tables</b>	<b>x</b>
<b>List of Figures</b>	<b>xi</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Job Referral Networks and the Determination of Earnings in Local Labor Markets</b>	<b>5</b>
2.1 Introduction . . . . .	5
2.2 A Model of Job Search with Referral Networks . . . . .	10
2.2.1 Job Offers Through Referrals . . . . .	12
2.2.2 Implications . . . . .	15
2.3 The Determination of Job Search Outcomes: Econometric Framework . . . . .	20
2.3.1 Econometric Model . . . . .	21
2.4 Data . . . . .	31
2.5 Stylized Facts . . . . .	37
2.5.1 Evidence of a job ladder . . . . .	38
2.5.2 Evidence of spatial correlation in $\psi$ . . . . .	42
2.6 Estimation Results . . . . .	48
2.7 Conclusion . . . . .	56
<b>3 Free to Move? A Network Analytic Approach for Learning the Limits to Job Mobility</b>	<b>58</b>
3.1 Introduction . . . . .	58
3.2 The Realized Mobility Network . . . . .	61
3.2.1 Basic Graph Theory . . . . .	61
3.2.2 Defining the Realized Mobility Network . . . . .	64
3.3 Related Literature . . . . .	66
3.4 Data . . . . .	68
3.5 Basic Topology of the Realized Mobility Network in the PSID . . . . .	71
3.5.1 Degree Distribution . . . . .	71
3.5.2 Clustering . . . . .	73
3.5.3 Distance . . . . .	74
3.5.4 Results . . . . .	75
3.6 Finding Mobility Clusters using Community Detection Across Workers . . . . .	82
3.6.1 Segment Structure . . . . .	86

3.6.2	Correlates of Worker Segment . . . . .	87
3.6.3	Wage Determination By Segment . . . . .	90
3.7	Finding Mobility Clusters by Maximum Likelihood Partition of Employers . . . . .	90
3.7.1	A Basic Model Of Job Mobility . . . . .	90
3.7.2	The Structure of Employer-Employer Links Predicted in the Model . . . . .	97
3.7.3	Maximizing the Log-Likelihood . . . . .	99
3.7.4	The Likelihood-Maximizing Partition . . . . .	101
3.7.5	Characteristics of the Segments . . . . .	102
3.8	Conclusion . . . . .	108
<b>4</b>	<b>Endogenous Mobility (with John M. Abowd)</b>	<b>112</b>
4.1	Introduction . . . . .	112
4.2	What is Endogenous Mobility and Does it Matter? . . . . .	115
4.2.1	Tests of Endogenous Mobility . . . . .	117
4.3	A Motivating Example of Endogenous Mobility with Comparative Advantage and Learning . . . . .	119
4.4	A General Model with Endogenous Mobility . . . . .	124
4.4.1	Model Setup . . . . .	125
4.5	Likelihood, Prior and Posterior Distributions . . . . .	127
4.5.1	Likelihood functions . . . . .	127
4.5.2	Prior distributions . . . . .	129
4.5.3	Posterior distributions . . . . .	129
4.6	Estimation Procedure . . . . .	131
4.7	Monte Carlo Estimates . . . . .	133
4.8	Conclusion . . . . .	136
<b>A</b>	<b>Job Referral Networks and the Determination of Earnings in Local Labor Markets</b>	<b>138</b>
A.1	Proof of Proposition 1 . . . . .	138
A.2	Proof of proposition 3 . . . . .	139
A.3	Proof of proposition 4 . . . . .	140
<b>B</b>	<b>Endogenous Mobility (with John M. Abowd)</b>	<b>142</b>
B.1	Details of the Monte Carlo Estimates . . . . .	142
B.1.1	Parameters for the Simulated Data . . . . .	142
B.1.2	Summary of the Simulated Data . . . . .	144
B.1.3	Posterior Mode . . . . .	146
B.2	A Proposed IV Estimator Based on Exploitation of the Full Realized Mobility Network . . . . .	147
B.2.1	The IV Estimator . . . . .	147
B.3	Formulas for Endogeneity Tests . . . . .	151
B.3.1	Test 1 . . . . .	151

B.3.2	Test 2 . . . . .	152
<b>Bibliography</b>		<b>155</b>

## LIST OF TABLES

2.1	List of Metropolitan Statistical Areas Used . . . . .	31
2.2	Descriptive Statistics . . . . .	35
2.3	Unconditional Transition Probabilities . . . . .	39
2.4	Sorting by log earnings and components from the AKM decomposition, $R^2$ method. . . . .	47
2.5	Regression Estimates . . . . .	50
2.6	Selection Correction Model Estimates . . . . .	52
2.7	Heterogeneous Referral Effects . . . . .	53
2.8	Quantile Regression Estimates . . . . .	55
3.1	Data Summary . . . . .	70
3.2	Basic Topology of the Realized Mobility Network in the PSID . .	76
3.3	Summary of Job Types in Coworker Segments . . . . .	86
3.4	Logit model for Coworker Segment . . . . .	88
3.5	Log Earnings Regression on Coworker Segments . . . . .	91
3.6	Probability of Transition Between Pseudoemployer Segments . .	101
3.7	Pseudoemployer Segment and Labor Market Earnings . . . . .	107
3.8	Factors Associated with Pseudoemployer Segment Assignment .	109

## LIST OF FIGURES

2.1	Cumulative probability of transition to each decile of the wage premium ( $\psi$ ) distribution, by decile of origin . . . . .	40
2.2	Probability of transition to each decile of the wage premium ( $\psi$ ) distribution, by decile of origin . . . . .	41
2.3	Spatial Autocorrelation Function: tract-level means . . . . .	43
2.4	Spatial Autocorrelation Function: block-level means . . . . .	44
3.1	Realized Employment Network: $t = 1$ . . . . .	63
3.2	Realized Employment Network: $t = 2$ . . . . .	64
3.3	Realized Employment Network: $t = 3$ . . . . .	64
3.4	Realized Mobility Network . . . . .	65
3.5	Degree distribution of the person nodes in the realized mobility network in the PSID . . . . .	77
3.6	Degree distribution of the pseudoemployer nodes in the realized mobility network in the PSID . . . . .	77
3.7	Degree distribution of the projection of the realized mobility network in the PSID sample onto workers . . . . .	78
3.8	Degree distribution of the projection of the realized mobility network in the PSID sample onto pseudo-employers . . . . .	80
3.9	Predicted probability of appearing in each of the six segments found by maximizing modularity of the worker projection. . . .	89
3.10	Predicted probability of appearing in each of the six segments found by maximizing modularity of the worker projection. . . .	89
3.11	Industry Distribution by Segment. Blue is Segment 0, Red is Segment 1. . . . .	103
3.12	Occupation Distribution by Segment. Blue is Segment 0, Red is Segment 1. . . . .	103
3.13	Occupation Distribution by Segment within FIRE. Blue is Segment 0, Red is Segment 1. . . . .	104
3.14	Occupation Distribution by Segment within Manufacturing. Blue is Segment 0, Red is Segment 1. . . . .	105
3.15	Occupation Distribution by Segment within the Service Sector. Blue is Segment 0, Red is Segment 1. . . . .	105
4.1	Test 1 of the exogenous mobility assumption: relationship between current period residual and past employer for workers who separate. . . . .	118
4.2	Test 1 of the exogenous mobility assumption: transition rates for workers with low earnings residuals . . . . .	119
4.3	Test 1 of the exogenous mobility assumption: transition rates for workers with earnings residuals in the fifth decile . . . . .	120

4.4	Test 1 of the exogenous mobility assumption: transition rates for workers with earnings residuals in the tenth decile . . . . .	121
4.5	Posterior distribution of parameters from the earnings equation based on draws from the Gibbs sampler. . . . .	135

## CHAPTER 1

### INTRODUCTION

Job mobility is a complex phenomenon, and network theory provides a novel and practical framework for dealing with this complexity in understanding how individuals move from job to job. The first chapter in this dissertation, *Job Referral Networks and the Determination of Earnings in Local Labor Markets*, studies the use of personal contacts to search for work is a ubiquitous and persistent feature of modern labor markets. How do social networks direct the flow of job information? What implications might they have for labor market efficiency, the spatial concentration of poverty, and income mobility? To shed new light on this topic, I model and empirically evaluate the role of neighborhood-level referral networks in job search. I identify positive and economically significant effects of the quality of jobs held by a worker's neighbors on the quality of his next job. These are the first results on direct local interactions in earnings determination estimated in the context of a job search model.

In a labor market characterized by search frictions, identical workers may be paid different amounts by different employers. Those looking for jobs seek out higher rents, and workers holding jobs prefer to share these rents with their friends and neighbors. To capture this process of social rent sharing, I extend an on-the-job search model with worker and employer heterogeneity to include social interactions in job search. In the model, individuals are more likely to sample a job with the same premium as their neighbors. Relative to a baseline model without contagion, social interactions generate excess correlation in outcomes between connected workers. The model also predicts that workers with better social contacts will experience better outcomes from job-to-job mobility.

To test these predictions, I use data on individual work histories from the Census Bureaus Longitudinal Employer-Household Dynamics (LEHD) Program matched to residential addresses. I estimate the job search model on estimates of firm-specific wage premia derived from a decomposition of log earnings into worker- and employer-specific heterogeneity components. The employer-specific wage premia conform to two stylized facts consistent with the model. First, there is evidence of a job ladder: workers tend to move into jobs with better wage premia. Second, there is significant spatial correlation of wage premia for jobs held by individual workers for detailed geographies. These stylized facts are novel, and provide evidence of the importance of sorting for the spatial distribution of earnings.

I identify the contribution to job search outcomes of the quality of local social networks from quasi-random assignment of workers to residential blocks within larger neighborhoods. This facilitates distinguishing neighborhood quality from network quality. Neighborhood quality affects search outcomes through a number of channels, for instance differential access to transportation. Workers residential location decisions are made on the basis of neighborhood quality, but they cannot sort perfectly by block. Thus, the variation in network quality within the neighborhood is exogenous. I measure ones local network quality from the distribution of employer-specific wage premia held by workers from the same residential block. My model predicts the excess spatial correlation found in employer wage premia at the block level beyond that found at the tract level. Both conventional and quantile regressions confirm that the relationship between network quality and job search outcomes is significant, economically meaningful, and conforms to the predictions of my enhanced search model.



The purpose of the research embodied in the second chapter is to place workers and employers into groups that are relatively homogeneous with respect to job mobility. Theory suggests that workers with interchangeable skills should have similar patterns of mobility across employers that use those skills in roughly the same way. Alternatively, in a model of labor market segmentation, workers in the secondary sector of the labor market cannot easily move into the primary sector in which access to jobs may be rationed. Both models imply that one can infer homogeneous groups of workers and jobs from patterns in longitudinal data on job mobility.

The key contribution of this paper is to apply the tools of network analysis to find these homogeneous groups of workers and jobs from patterns of job mobility. Job-to-job mobility is modeled as a realized mobility network connecting workers to employers. The empirical component of the paper uses data from the PSID to construct a realized mobility network connecting workers to the detailed industry-occupation pairs in which they are employed. Assuming workers are more likely to change jobs within a segment than to exit yields a likelihood function over all possible partitions of the network. Although this problem has high dimension, it is possible to maximize the likelihood function by simulated annealing. The resulting estimates group industry-occupation pairs into two groups with distinct employment patterns, and compensation profiles.

Separating jobs solely on the basis of mobility patterns reveals a small sector in which earnings determination is highly correlated with education and minimally correlated with demographic characteristics relative to the larger sector. The smaller sector is made up of professional and managerial occupations in the service and finance sectors. By contrast, the larger sector is made up of clerical,

service, and routine labor. Using an alternative measure of network clustering, I find evidence that the labor market contains at least five large economically and demographically-distinct groups of workers. The same patterns are replicated, but with some refinement. The largest two sectors are split between service and manufacturing jobs at all occupational levels, with three smaller sectors centered on specific occupational groups.

The final chapter, *Endogenous Mobility*, is joint with John Abowd. We propose and implement a method for bounding the amount of endogeneity bias in estimates of person- and employer-specific earnings heterogeneity. Unbiased and efficient estimates of these quantities are needed for applications in labor economics, empirical industrial organization, and macroeconomics. We instrument for observed work histories with simulations drawn from a model for the evolution of the realized mobility network that connects workers and employers over time. To demonstrate this technique, we specify and estimate a model in which both earnings, job separations, and job assignments are fully endogenous. Using simulated data, we estimate the posterior distribution of the parameters governing wage determination. We find that we are able to detect and correct for endogenous mobility bias in the simulated data.

CHAPTER 2

**JOB REFERRAL NETWORKS AND THE DETERMINATION OF  
EARNINGS IN LOCAL LABOR MARKETS**

## **2.1 Introduction**

In this paper, I study a previously unexplored connection between two features of the U.S. labor market. The first is that who you know affects where you work. The second is that where you work affects how much you are paid. Getting job information from friends and neighbors is a common strategy, and apparently a productive one: between 30 and 60 percent of new jobs are found through personal contacts (Bewley 1999). Why is referral use so prominent? What does this prominence mean for the way the labor markets operate? These remain open questions. One role for referral networks in job search is in helping workers locate information about particularly attractive job opportunities. Two workers can receive different pay simply because they work in different firms (Abowd et al. 1999). If workers share information about these pay differentials with their friends and neighbors, then who you know can affect how much you are paid.

The goal of this paper is to identify the role of local referral networks in the assignment of employer-specific pay differentials. In doing so, I provide the first direct evidence of referral effects and neighborhood interactions in earnings determination. I find that workers engaged in on-the-job search receive a positive and significant fraction of their job offers through local interactions. Furthermore, workers who change jobs receive offers with higher pay differentials when workers in their local referral networks are earning higher differentials. These results are robust to various specifications that attempt to address sorting

and to correct for sample selection on the quality of one's current job. The magnitude of the effect is very similar to self-reported levels of referral use among employed workers found in survey data (Ioannides and Loury 2004). My empirical approach relies on workers being likely to interact with their residential neighbors in searching for better jobs. To check the validity of my approach, I extend the model to allow for differences in the productivity of local referral networks between native and non-native workers. I find that the magnitude of the effect of local interactions on job quality are almost twice as strong for non-native workers as for natives, consistent with previous research showing that non-native workers are more likely to use referrals to find work.

My research makes several important contributions to the literature on the role of local social interactions in labor market outcomes. This is the first paper to directly identify and estimate local spillovers in earnings determination. Previous research on neighborhood effects has either focused on neighborhood-level effects in social behaviors correlated with income, or on how neighborhood characteristics affect labor market outcomes (see Ioannides and Loury (2004) for an extensive survey). I am able to make progress because I can separate the part of earnings associated with job assignment from the part due to an individual worker's portable skills. I also have a clean strategy for identifying social interactions in job assignments and earnings by exploiting variation in network quality within neighborhoods. This strategy is derived from Bayer et al. (2008) who use it to identify local interactions in job location. They do not focus on direct spillovers in earnings determination, and do not relate their findings to a formal model of job search. Topa (2001) and Conley and Topa (2007) do use a formal job search model to estimate local interactions in unemployment in a formal job search model, but lack a clean identification strategy. In addi-

tion to its primary contributions, this paper is the first to verify that employer-specific wage premia have a mobility-related structure that is consistent with a job search model. It is also the first to document the spatial structure in the distribution of earnings components using employer-employee matched data for the United States.

To obtain my results, I develop and empirically implement a model of job search in which workers use local referrals to share information about pay differentials. In order to outline the model and empirical work, I reserve the term ‘quality’ to refer to the pay differential on a particular job in a shorthand way. In this paper, I focus just on quality in terms of employer-specific pay differentials, finding that this dimension of quality matters theoretically and empirically.

I start by modelling a job search process in which information about job quality passes through referral networks. Consistent with empirical evidence and a range of job search models, employers are distinguished by idiosyncratic pay differentials. These pay differentials are non-market rents that accrue to the workers who find them. A worker searching for a job can use his referral network to try to find these rents. Likewise, a worker who knows about an employer with particularly attractive terms will share that information with his friends and neighbors. In the model, I allow workers to search for better jobs through direct contact with employers or by sampling offers of the same quality as one of their employed social network contacts. The distribution of job offers has a simple form in which the average quality of a worker’s job offer depends directly on the the average quality of his neighbors’ jobs. I derive an econometric model for the offer function along with implications for the conditional mean and quantiles of the observed job quality distribution.

Then, to identify local referral effects in job quality, I use employer-employee matched data from the Longitudinal Employer-Household Dynamics (LEHD) Program. The empirical work has two components. First, I obtain measures of the quality of jobs held by all private sector, non-farm workers. I take these from a decomposition of log earnings into components associated with individual and employer heterogeneity as described in Abowd et al. (2002). I then link the job quality estimates to the exact residential block for workers who lived in one of 30 large Metropolitan Statistical Areas (MSAs) in 2002-2003. Using these data, I identify local interactions in job quality by exploiting variation in the average quality of jobs within a relatively small reference group of neighboring blocks.

The empirical model is related to the conventional neighborhood effects specification where an individual's expected behavior depends on the average behavior of his neighbors. The model therefore raises the same identification issues first noted by Manski (1993) and summarized by Blume and Durlauf (2005). I avoid Manski's reflection problem by exploiting the time-dimension in my data to construct measures of local referral quality that are predetermined at the time of an observed job transition. More problematic is the difficulty of separating the effects of local referral effects from other economic processes, including residential sorting, that might cause job search outcomes to be spatially correlated.

Adapting the research design of Bayer et al. (2008), I identify the contribution to job search outcomes of the quality of local social networks from quasi-random assignment of workers to residential blocks within larger neighborhoods. This facilitates distinguishing neighborhood quality from network quality. Neighborhood quality affects search outcomes through a number of channels, for in-

stance differential access to transportation. Workers residential location decisions are made on the basis of neighborhood quality, but they cannot sort perfectly by block. Thus, the variation in network quality within the neighborhood is exogenous. I measure ones local network quality from the distribution of employer-specific wage premia held by workers from the same residential block. My model predicts the excess spatial correlation found in employer wage premia at the block level beyond that found at the tract level. Both conventional and quantile regressions confirm that the relationship between network quality and job search outcomes is significant, economically meaningful, and conforms to the predictions of my enhanced search model.

The estimation results are driven by two stylized features of the distribution of estimated employer-specific wage premia. First, there is a ‘job ladder’ in the sense that workers who change jobs are more likely to move from lower- to higher-quality jobs (Figure 2.2). Second, job quality is spatially correlated at the level of the Census block (Figure 2.4). I show that these features are characteristic of the on-the-job search model with local referral networks, and then show that they hold in the estimated distribution of employer wage premia. My analysis of the spatial correlation of earnings, human capital, and employer characteristics is among the most geographically detailed of its type for U.S. cities. My results also confirm that much of the observed sorting in earnings is correlated with sorting on observable and unobservable human capital characteristics (Combes et al. 2008; Conley and Topa 2002a). These results are relevant to those studying residential sorting by earnings, human capital characteristics, and employer characteristics in urban labor markets.

## 2.2 A Model of Job Search with Referral Networks

I model on-the-job search with social interactions in the transmission of information about new job opportunities. Different employers offer different pay to the same worker, but workers do not know the size of the wage premium offered by any particular employer. They must engage in a process of search to collect information about new jobs. I allow for the possibility that the productivity of the search process may depend on individual characteristics, neighborhood quality, and the quality of the jobs held by people in one's referral network.

This model of wage-setting is motivated by the empirical finding that employer specific heterogeneity explains a large portion of the dispersion in log earnings (Abowd et al. 1999; 2002). This is consistent with a primary theoretical result of job search models, which show that information imperfections lead labor markets to fail to eliminate all idiosyncratic differences in pay between employers (Rogerson et al. 2005). The wage function given here could arise in a matching model with worker and employer heterogeneity in production with surplus sharing when there is no wage renegotiation (Postel-Vinay and Robin 2002).

Time evolves continuously and the observed data are snapshots taken at discrete intervals. To clarify the presentation, I denote a model variable, say earnings of the  $i^{th}$  worker, evolving in continuous time, as  $y(i, t)$ . The data observed from this process are denoted as  $y_{i1}, y_{i2}, \dots, y_{iT}$  where each  $y_{i\tau}$  is an observation on this process at time  $t = \tau$ . For the remainder of this section, I use notation indicating the continuous time evolution of the model.



The model is populated by a finite group of workers and a continuum of employers. Let  $i \in \{1 \dots I\}$  index workers and  $j$  index employers. Workers are heterogeneous in the characteristics that affect productivity and pay. Let  $e(t; i)$  denote the stock of human capital characteristics held by worker  $i$  at time  $t$ . Different employers compensate workers differently. Let  $p_j > 0$  be the idiosyncratic component of employer pay. The earnings function,  $y(e(t; i), p_j)$  satisfies log-separability. That is,

$$\ln y(e(t; i), p_j) = \ln y_1(e(t; i)) + \psi_j,$$

where  $\psi_j = \ln y_2(p_j)$ .  $\psi_j$  is the log-wage premium paid by employer  $j$ .

The model specifies the continuous time evolution of the  $I \times 1$  vector of assignments of workers to employers with different wage premia,  $\Psi(t)$ . Workers are infinitely lived and can be either employed or unemployed. They search for jobs paying higher premia. The arrival of job offers is a Poisson process with unemployed workers receiving new job information at Poisson rate  $\lambda_0$ . Employed workers receive job information at rate  $\lambda_1$ . Jobs can end due to exogenous productivity shocks that occur at rate  $\delta$ . I assume that these contact and separation rates are exogenous and common across workers. The only decision that workers make is whether to accept a job when they receive an offer. I assume that workers receive utility in unemployment equivalent to getting a job with wage premium  $p_b$ .

When a worker receives a job offer, it is sampled from an employer offering the log wage premium  $\psi$  with probability  $f(\psi; i, t)$ . As the notation indicates, the sampling distribution differs across workers and can change over time. This distribution is a mixture of a formal market offer distribution, denoted  $g(\psi; i, t)$  and the distribution of job offers of one's social contacts, denoted  $h(\psi; i, t)$ .

The parameter  $a$  measures the strength of social interactions relative to formal channels in delivering new job offers. It is the object of primary interest in the empirical analysis. I specify the relationship between formal search and referral networks as a simple mixture. A worker samples an offer,  $\psi$ , from either the distribution of offers in the formal market,  $g(\psi; i, t)$ , or from the distribution of offers that come through his referral network,  $h(\psi; i, t)$ . Thus conditional on receiving an offer, the worker draws its type from the distribution

$$f(\psi; i, t) = ag(\psi; i, t) + (1 - a)h(\psi; i, t). \quad (2.1)$$

In setting up the model, I maintain that  $a$  is identical across workers. In the empirical work, I estimate the model under this restriction, but also allow for heterogeneity in  $a$  on observable characteristics. For this to be a model of job information transmission through referrals, it is necessary to explicitly specify the relationship between social structure and the distribution of job offers. I turn to this task next.

### 2.2.1 Job Offers Through Referrals

Social interactions in job search follow a contact process in which individuals receive information about job opportunities from their neighbors. The transmission of this job information is stylized as a contagion process from epidemiology. When searching, workers either sample from a fixed ‘formal market’ offer distribution, or may sample an offer of the same type as one of their employed friends. The formal offer distribution describes the availability of jobs received when applying directly to employers, answering ads or knocking on doors. The informal offer distribution describes the probability of receiving a job of a particular type conditional on the number of your social contacts who already hold

that type of job. Here, the types of jobs held by ones neighbors are ‘contagious’ in the sense that their social network partners are at increased risk to get an offer for the same type of job as one they already hold. This captures the intuition that personal referrals are explicitly used to find and share information about particularly attractive, employer specific wage premia. Conley and Topa (2007) assumes a similar contact process, but for the case of transitions into employment, whereas I consider transitions across employers with different productive characteristics.

Let  $W$  be an  $I \times I$  stochastic matrix whose  $(ji)^{th}$  entry measures the probability that job information received by  $i$  through a referral originated with worker  $j$ . In the empirical work, I relax the assumption that social structure is exogenous since I allow for the possibility that people sort into neighborhoods on the basis of unobservable characteristics that might be correlated with their job search outcome. I consider the class of social interactions models for which the distribution of offers received through referrals satisfies

$$E_h(\psi|W, i, \Psi(t)) = (w^i)^T \Psi(t), \quad (2.2)$$

where  $w^i$  is the  $i^{th}$  column of  $W$ . This specification is consistent with models of information transmission where proximity to workers with better wage premia results in better expected offers. One such model is a contagion process where the probability of receiving an offer with log wage premium  $\psi$  is increasing in proximity to workers already holding jobs paying that premium:

$$h(\psi; i, t) = (w^i)^T \mathbb{1}(\Psi(t) = \psi). \quad (2.3)$$

This model captures the intuition that referrals are used to share information about particularly attractive wage premia. In the empirical work, I identify the

effects of social networks by the quasi-random variation in the residential location choices of individual workers. Since this variation facilitates the identification of local neighborhood interactions, I will specify  $W$  in terms of residential proximity.

The construction of  $h$  in equation 2.3 embeds the assumptions that there are no demand-side constraints that affect the distribution of offers through the referral network. This is consistent with the partial equilibrium nature of the model. Second, and more crucial, is the assumption that the probability that  $i$  receives an offer  $\psi$  through referral,  $h(\psi; i, t)$  is independent of the offer received by another worker  $k$  at  $t$ . In other words,  $i$  and  $k$  are not competitors for the same scarce piece of job information. This assumption is a key feature of the contagion approach, and differs from related models that focus on the routing of job information across social networks in partial or general equilibrium search and matching models (Calvo-Armengol and Jackson, 2004; Calvo-Armengol and Zenou 2005, Wahba and Zenou 2005). Their models emphasize congestion effects in the transmission of job information alongside contagion effects. Congestion occurs when many workers in a social network are competing for the same job information. As Wahba and Zenou (2005) have shown, network congestion effects lead to empirically verified non-linearities in the use and effects of social contacts to find work.

In this paper, I abstract from congestion effects to focus on identifying the effect of local network quality on job search outcomes. This abstraction eliminates dependencies between worker's outcomes in the instantaneous cross-section. More plainly, taking network quality as given, the job offers received by any worker are independent of those received by any other worker. This assump-

tion is approximately correct if the congestion effect is trivially small relative to the contagion effect. Modeling social interactions in job search as a contagion process allows independence in individual job search outcomes.

This approach is an extension of Mortensen and Vishwanath (1994). They develop a general equilibrium labor market model with on-the-job search. Like my model, workers can either sample an offer from the formal offer distribution, or sample directly from the distribution of realized job offers. This is equivalent to my model in the case where there is no heterogeneity in the underlying referral network and all workers are equally likely to be sampled for job information. Calvo-Armengol and Jackson (2007) also have a partial equilibrium model of job search, but allow for a more general specification of transmission of job information. Their model yields clean theoretical results, but does not lend itself as easily to empirical evaluation. Some of the papers already discussed, along with the work of Fontaine (2007) and Cahuc and Fontaine (2002), use a different kind of contact process to stylize the social transmission of job information. These ‘urn-ball’ models facilitate evaluation of equilibrium dynamics in frictional search and matching models, but do not lend themselves to estimation of social interaction effects in job search outcomes from data on individual job histories.

### **2.2.2 Implications**

The search model with these simple behavioral rules regarding job mobility yields a continuous-time Markov process over assignments of workers to types of jobs. In the case where the social interaction parameter  $(1 - a) \equiv 0$ , there are

no referral network effects, and the observed job transitions are *i.i.d.* samples from the same (conditional) Markov process. When  $a \neq 1$  there are spillovers leading to correlations in the state vector. It is conceptually straightforward to define a transition kernel,  $Q(Z, W)$  for the evolution of  $\Psi(t)$  from the primitives of the mobility model,  $\lambda, \delta, a, h$  and  $g$ . The notation reflects the dependence of the kernel on a matrix of observable worker characteristics,  $Z$  and the social distance matrix,  $W$ . The full mobility model has the form:

$$Y(t) = \ln y_1(E(t)) + \Psi(t) + \varepsilon(t) \quad (2.4)$$

$$Pr(\Psi(t)|\Psi(t - \Delta)) = \exp(Q(Z, W)\Delta)\Psi(t - \Delta), \quad (2.5)$$

where  $Y(t)$  is a vector of observed log earnings,  $E(t)$  is an  $I \times 1$  matrix of time-varying human capital characteristics, and  $\varepsilon(t)$  is a vector of errors. The term  $\exp(Q(Z, W)\Delta)$  refers to the matrix exponential. The model delivers simple predictions for the stationary distribution of  $\Psi(t)$ .

First, I derive some basic properties of a job search model with social interactions in job offers. These suggest certain stylized features that should be observed in the data. Job-to-job transitions will involve moves to employers paying higher wage premia. Second, job mobility is a Markov process whose steady state distribution generates excess correlation between workers in the same referral network. This is related to a similar finding in Calvo-Armengol and Jackson (2007) I argue on the basis of simulation results that correlations in the employer-specific wage premia are stronger for workers who are socially closer according to  $W$ .

That workers always accept a job paying a higher premium follows from the assumptions that they like money and that the evolution of their portable skills,  $e(t; i)$ , is independent of job assignment. The latter assumption may not hold

if workers choose jobs both for their wage premia and also to optimize wage growth associated with experience in particular sector. It allows me to focus on the role of wage premia in job assignment and job mobility. It is probably not a bad approximation for workers who supply labor in jobs where there is little human capital specificity, and also for workers who have already selected a career and are changing jobs within their chosen field to maximize earnings (Neal, 1999). My main results are based on estimates of the model for all workers, but to acknowledge the preceding argument, I also allow for heterogeneity in the parameter  $a$  to accommodate the possibility that the model may more accurately describe certain groups of workers than others. To foreshadow the results, I find that my estimates of local interactions in job search are much stronger for non-native than for native workers.

Even though it is a standard feature of on-the-job search models, to clarify the restrictions on the present model needed to establish the result that workers move to jobs paying higher premia, I restate it formally as a proposition:

*Proposition 1* In the job search model described above, assume workers are expected wealth maximizers and  $e(t; i)$  is independent of work history. Further, assume workers are myopic about the evolution of the offer distribution. Then employed workers will always accept an offer of a job paying a higher wage premium. In addition, unemployed workers follow a reservation strategy.

**Proof.** See Appendix A. ■

The previous result is true for most models of on-the-job search. The next result is specific to a model with on-the-job search with social transmission of job information. It simply states that the correlation in job assignment for socially

connected workers is positive, and increasing in their social proximity according to  $W$ . If this is not true, then it is not clear that there is any empirical content to the social interaction model. Specifically, I claim the following:

*Proposition 2* The stationary distribution of  $\Psi$  is such that

$$W_{i''} > W_{i'} \text{ implies } \text{cov}(\psi_i, \psi_{i'}) > \text{cov}(\psi_i, \psi_{i''})$$

That is, correlations in employer-specific wage premia are stronger for workers who are socially closer according to  $W$

Supporting Proposition 2 are results from a simulation of the contagion model. Workers reside in a set of contiguous blocks that are arranged in a circle, so that each block is adjacent to two other blocks. Every block holds 100 workers, and they are socially connected to workers residing on the same block, and also to workers residing on adjacent blocks. The social proximity to workers on the same block is greater than the proximity to workers on adjacent blocks. I assume zero social proximity between workers who do not live on the same or adjacent blocks. Specifically, workers are expected to have 15 social contacts out of 100 on their block and anywhere from 0 to 10 out of 100 on each adjacent block. There are only two wage premium levels in the simulation. Without social interactions, that is when the number of social contacts on adjacent blocks is zero, the spatial correlation in the fraction of workers earning high wage premia between adjacent blocks is 0.0065. As the number of contacts rises from zero to ten, the the spatial correlation rises to 0.85. Thus, in an economy with stronger social interactions among adjacent blocks, we see increased spatial correlation, consistent with the theoretical prediction.

Once these two results are verified in the data, the goal is to check whether



the relationship between referral network quality and job search outcomes exists and conforms to the predictions of the model. A job-to-job switch is an observation from a stochastic process whose mean is  $E(\psi|\psi > \psi_0, Z, W, \Psi^0)$  where  $\psi_0$  is the log wage premium on the worker's current job and  $\Psi^0$  is the vector of log wage premia held by all workers at the time of the transition. The following proposition shows that an increase in network quality will increase the mean of the truncated offer distribution:

*Proposition 3* If the distribution of offers received through referral,  $h$ , is log concave and  $|E_g(\psi|\psi > \psi_0, Z, W, \Psi^0) - E_h(\psi|\psi > \psi_0, Z, W, \Psi^0)|$  is small then

$$\frac{\partial \mu_{f^*}(\psi_0)}{\partial \mu_h} > 0$$

where  $\mu_{f^*}(\psi_0) = E(\psi|\psi > \psi_0, Z, W, \Psi^0)$  and  $\mu_h = E_h(\psi|Z, W, \Psi^0)$

**Proof.** See Appendix A. ■

The requirement that  $|E_g(\psi|Z, W, \Psi^0) - E_h(\psi|Z, W, \Psi^0)|$  is small may be restrictive, but is probably satisfied in practice. It says that the distribution of acceptable offers from referrals is not too different from the distribution of acceptable offers from formal search. The jobs available through the referral network should generally be fairly close to the distribution of offers that workers would receive through formal search, including those features of job search productivity that are correlated across individuals.

The job search model also yields predictions on the quantiles of the truncated offer distribution. I evaluate these in the empirical work as additional checks of the validity of the job search model

*Proposition 4* If the cumulative distribution function of the wage premium offer distribution,  $F(\psi)$ , is log concave, twice continuously differentiable, and its

density function symmetric, then (i) an increase in  $\psi_0$  has a monotonically decreasing effect on quantiles of the  $\psi$  distribution, (ii) anything that increases the mean of the offer distribution has an increasing effect on quantiles of the  $\psi$  distribution.

**Proof.** See Appendix A. ■

In the search model, increasing  $\psi_0$  affects outcomes by increasing the reservation offer that triggers mobility. Intuitively, increasing  $\psi_0$  will have a larger impact on the distribution of acceptable offers close to the truncation point than those further away. The condition of proposition 4, that the offer distribution is log concave with a symmetric density, is satisfied by the normal distribution, the uniform distribution, and the double exponential.

### 2.3 The Determination of Job Search Outcomes: Econometric Framework

I estimate the social interaction parameter using data on observed job-to-job transitions that have been linked to data on workers' blocks of residence. I follow a two-stage estimation procedure. The first stage consists of estimating employer-specific wage premia from a complete history of matched employer-employee data. The second stage treats these estimates as data in estimating the offer function for employed workers. This procedure is consistent under the assumption that the wage error  $\varepsilon(t)$  from equation 2.4 is orthogonal to the mobility process in equation 2.5. This assumption is explicit in my search model, and conforms to the standard exogenous mobility assumption for consistent es-

timization of the parameters of the log-linear earnings decomposition (Abowd et al. 1999).

To assuage concerns about the validity of these assumptions, I test that the distribution of estimated wage premia,  $\hat{\psi}$ , conforms to the Propositions 1 and 2. I show that  $\hat{\psi}$  is consistent with a noisy job ladder form of job-to-job mobility. A worker is more likely to move to a job with an employer paying a higher log wage premium than his current employer. Second, the data exhibit spatial correlation in  $\hat{\psi}$  at the level of the Census block, consistent with contagious search when social interactions occur among residential neighbors.

Having established these results, I proceed to estimate the offer function using data on workers who make direct job-to-job transitions. The following discussion describes the econometric model for the offer function. In what follows, I completely specify the wage and offer function, calibrate the referral network,  $W$ , and describe my identification strategy. I conclude by describing the full econometric model that accounts for sample selection associated with the job search model.

### 2.3.1 Econometric Model

I make several assumptions to facilitate bringing the model to the data. Recall that observed earnings are denoted by  $y_{it}$ , and specify the earnings determination process so that

$$y_{it} = \gamma e_{i,t} p_{J(i,t)} \quad (2.6)$$

$$\ln y_{it} = \ln \gamma + \ln e_{i,t} + \ln p_{J(i,t)}. \quad (2.7)$$

$J(i, t) = j$  where  $j$  is the employer of  $i$  at time  $t$ . Human capital depends on observable time-varying inputs,  $X_{it}$  and observable and potentially unobservable correlates of ability,  $\theta_i$ , so that

$$e_{it} = \exp(X_{it}\beta + \theta_i).$$

Since  $\psi_j = \ln p_j$  the final expression for log earnings is:

$$\ln y_{it} = \alpha + X_{it}\beta + \theta_i + \psi_{J(i,t)} + \varepsilon_{it}. \quad (2.8)$$

The model allows arbitrary heterogeneity in the formal and informal offer distributions:

$$f_{it}(\psi) = ag_{it}(\psi) + (1 - a)h_{it}(\psi). \quad (2.9)$$

I assume that this heterogeneity is fully captured by observable worker characteristics,  $Z_i$ , the vector describing  $i$ 's referral network,  $w^i$ , and the log wage premia held by workers at the time of the transition. The latter quantity is the data analogue to  $\Psi(t)$ , denoted  $\Psi^t$ , where the  $i^{th}$  entry is  $\psi_{J(i,t)}$ , the log wage premium paid by employer  $j = J(i, t)$ . The offer distribution is:

$$f(\psi|Z_i, w^i, \Psi^t) = ag(\psi|Z_i, w^i, \Psi^t) + (1 - a)h(\psi|Z_i, w^i, \Psi^t). \quad (2.10)$$

It is a simple formality to express a realized offer,  $\psi_{i,t}^*$  in terms of the means of the formal and informal distributions,  $g$  and  $h$ , and deviations from those means.

$$\psi_{i,t}^* = a(E_g(\psi|Z_i, w^i, \Psi^t) + \eta_{i,t}^g) + (1 - a)(E_h(\psi|Z_i, w^i, \Psi^t) + \eta_{i,t}^h) \quad (2.11)$$

$$= aE_g(\psi|Z_i, w^i, \Psi^t) + (1 - a)E_h(\psi|Z_i, w^i, \Psi^t) + \eta_{i,t}, \quad (2.12)$$

where  $\eta_{i,t} = a\eta_{i,t}^g + (1 - a)\eta_{i,t}^h$ . Restrictions on the sources of observable variation and the error processes clarify the essential identification problem and provide a template for implementing the model empirically. The model specifies the

mean of the informal offer distribution in equation (2.2), which is implemented empirically as:

$$E_h(\psi|w^i, \Psi^t) = (w^i)^T \Psi^t.$$

Furthermore, the expected offer from formal search is conditionally independent of the social network and the assignments of other workers to jobs. So we have

$$E(\psi_{it}^*|Z_i, w^i, \Psi^t) = aE_g(\psi_{it}^*|Z_i) + (1-a)(w^i)^T \Psi^t + aE(\eta_{i,t}^g|w^i, \Psi^t) + (1-a)E(\eta_{i,t}^h|Z_i). \quad (2.13)$$

These restrictions imply that the mean offer received through formal channels depends on individual characteristics but not on the the structure of the referral network or current set of jobs held by other workers. The mean offer received through the referral network depends only on network position and the quality of jobs held by other workers. Also, the expected deviation of the formal offer from the mean does not depend on individual characteristics, but may be correlated with the social structure. These issues highlight the inherent identification problems.  $a$  is identified if

$$E(\eta_{i,t}|W_i, \Psi^t) = aE(\eta_{i,t}^g|w^i, \Psi^t) + (1-a)E(\eta_{i,t}^h|Z_i) = 0. \quad (2.14)$$

I abstract here from sample selection generated by the fact that only acceptable job offers are observed, but return to it below. Assuming all job offers are observed, then as long as there is variation in  $E_h(\psi_{it}^*|W_i, \Psi^t)$  that is uncorrelated with  $E_g(\psi_{it}^*|Z_i)$ ,  $a$  is identified under the model. Even when there is sample selection, this variation still facilitates identification as long as there is additional exogenous variation driving the selection process. Under certain circumstances, the selection process may itself assist in identification. See (Blume and Durlauf 2005) and (Brock and Durlauf 2001) for details.

The assumption that  $E(\eta_{i,t}^h | Z_i) = 0$  is reasonable. If the social interaction process has been properly specified, then the influence of one's own characteristics on the arrival of offers through referral should already be included through  $w^i$ . The assumption that  $E(\eta_{i,t}^s | w^i, \Psi^t) = 0$  may be too strict. It attributes all correlation in outcomes for individuals that are socially proximate according to  $W$  to the effect of referral networks. However, social proximity may be correlated with factors that influence formal job search. First, workers may sort on the basis of unobservable characteristics that determine the efficiency of formal search. There may also be variables determining  $W$  that are correlated with search outcomes. For instance, workers in the same residential neighborhood are closer according to  $W$ , but they also experience correlated search outcomes, for instance because of differential access to transportation.

The empirical challenge is to distinguish between these different causes of correlation in job offers. I am able to distinguish the effects of spatially correlated factors related to the efficiency of job search from the effect of referral networks because I observe variation in network quality at very high levels of geographic detail. Workers living in the same neighborhood confront minor variations in the quality of their referral networks since residential proximity is a determinant of social proximity. Within a neighborhood, spatially correlated factors affecting search are identical across workers. Therefore, the effects of referral networks are identified by controlling for unobserved effects driving search outcomes at the neighborhood level. The network effect is identified from the within-neighborhood variation in network quality.

I make a parametric assumption that the conditional mean of the formal offer

distribution is linear in observable worker characteristics.

$$E_g(\psi_{it}^*|Z_i) = Z_i\tilde{\Pi}. \quad (2.15)$$

This allows for the possibility that some groups of workers have better luck in finding jobs paying high wage premia, perhaps because of discrimination against particular groups, or because members of the same demographic group tend to search for jobs in the same way. In the empirical work, I check for robustness of my results to variations in this modeling assumption. The offer function is therefore given by

$$\psi_{i,t}^* = Z_i\Pi + \gamma(w^i)^T\Psi^t + \eta_{i,t}, \quad (2.16)$$

where  $\gamma = (1 - a)$  and  $\Pi = a\tilde{\Pi}$ .

### Definition of the Reference Group

Empirical implementation requires a definition of the referral network. I assume that  $W$  puts equal weight on all people residing on the same block, and no weight elsewhere. That is,  $(w^i)^T\Psi^t = \bar{\psi}_{b(i)t}$  where  $b(i)$  indicates the block of residence for worker  $i$ , and  $\bar{\psi}_{b(i)t}$  is the average wage premium in jobs held by workers at time  $t$ .

$$\psi_{i,t}^* = Z_i\Pi + \gamma\bar{\psi}_{b(i)t} + \eta_{i,t}. \quad (2.17)$$

This definition of  $W$  conforms to many existing studies of neighborhood effects in labor market outcomes (Topa 2001; Weinberg et al. 2004; Bertrand et al. 2000; Case and Katz 1991; Bayer et al. 2008). Focusing on local residential neighborhoods exploits the variation in the data. At the same time, it is clear that transmission of job information described in the model is by no means restricted to conversations among people living on the same block. Conley and Topa (2002a)

demonstrate that correlations in employment status are well explained by occupational distance and racial or ethnic distance and that social distance metrics combining multiple dimensions of social proximity are appropriate. As in the analysis by Bayer et al. (2008), as long as it is correct that social interactions of the type hypothesized here are stronger among near neighbors, estimates based on this particular calibration put a lower bound on the true extent of referral network activity.

Alternative calibrations that allow for social distance to depend on race, ethnicity, nativity and their interactions are feasible. This flexibility is an appropriate response to model uncertainty about social structure. Identification of referral effects with an arbitrary matrix  $W$  using a design similar to the one in this paper may be possible under extensions of the identification strategy proposed by Bramoullé et al. (2009), though these will be computationally intensive. Furthermore, resolving model uncertainty about  $W$  awaits tools for separately identifying social interaction effects from variation in social network structure  $W$ . Manski (1993) warns that observed data cannot be used to infer reference group structure, which if true would mean it is not possible to simultaneously learn  $W$  and the effect of  $W$  from the data.

Naturally, proper care must be taken in implementing this empirically to make sure that  $\bar{\psi}_{b(i)t}$  is computed in such a way that the worker's initial job and next job are not included, and that only jobs that were already in progress at the time the worker made his transition are included. I discuss this in detail when describing the estimation procedure.



## Identification

The task of identifying social interactions in earnings determination is a special case of a general identification problem documented by Manski (1993) and elaborated in Brock and Durlauf (2001). Blume and Durlauf (2005) provide a concise introduction to this literature with a useful discussion of the kinds of data and models that can be used to identify social interactions. My paper is able to make progress in identifying local interactions in job search for two reasons. First, I use the time dimension of the LEHD data to measure the quality of jobs in a worker's referral network prior to him changing jobs. The search model describes a sequential process of interactions that exploits the longitudinal structure of the data. The conditional independence between past and future job assignments breaks the reflection problem described by Manski (1993). Conley and Udry (2010) also use the time sequencing of information transmission to identify the effect of social learning by farmers in Ghana about new agricultural practices.

Second, the large sample size and fine detail of the residential address information mean that correlated effects driving job search outcomes can be separately identified from referral network quality under mild assumptions. Following Bayer et al. (2008), I assume all economically plausible factors affecting formal job search are homogeneous within pre-defined reference groups of geographically contiguous Census blocks. Referral network effects are then identified by block-level variation in network quality within these homogeneous groups. Many economic processes generate spatially or temporally correlated outcomes from formal search. One particularly problematic process is the residential sorting of workers in terms of latent characteristics that affect job search.

Also, neighborhoods differ in their proximity to jobs with particular characteristics so that workers in those areas have correlated search outcomes simply due to proximity.

I assume that each of these economic processes generates correlation in search outcomes for workers living in the same reference group correlation, but there is no excess correlation at the block level. The economic rationale for this is as follows: Residential sorting is imperfect. Thinness of the residential real estate market means that workers can choose the neighborhood in which they live, but generally not a specific block. Similarly, employers may prefer to hire workers from a certain part of the city, but it is unlikely that they have strict preferences for workers from specific blocks within the same neighborhood. Finally, in urban areas, transportation access is similar for workers residing in the same neighborhood.

For the empirical work, I use Census block groups as the reference group. Block groups are a convenient choice for several reasons. They are the lowest level of geography above the block for which the Census Bureau releases data, and are structured to collect relatively homogeneous, geographically contiguous blocks that do not cross tract boundaries. Census block groups have an optimal size of 1,500 occupants, but there is considerable variance in the number of people in any particular block group.

To operationalize the model and identifying assumptions, I use a cross section of data on workers residing in 30 large MSAs who are observed to make direct job-to-job transitions. I exploit the longitudinal structure of the work history data to construct measures of local referral network quality that are predetermined at the date of the worker's own transition. The final specification of

the offer function is:

$$\psi_i^* = Z_i\Pi + \gamma\bar{\psi}_{b(i)0} + \zeta_{G(b(i))} + \eta_i. \quad (2.18)$$

where  $\bar{\psi}_{b(i)0}$  is the within-block average wage premium across all employed workers whose jobs were already in progress before the quarter in which  $i$  makes a transition, and that remained in progress in the quarter after.  $\zeta_{G(b(i))}$  is a reference group effect where  $G(b(i))$  denotes the reference group within which  $b$  belongs.

### A Formal Econometric Model with Sample Selection

The observed wage premium distribution is truncated due to the job-ladder behavior of workers. Without measurement error or non-pecuniary benefits associated with employment at a particular firm, an already employed worker changes jobs only on receipt of an offer for a job with a higher premium. The wage premia estimated in the first stage show workers will move to jobs with lower premia. This feature of the data is related to the finding in Nagypal (2005) that the rate of job-to-job transitions is not consistent with the strong job ladder model. They are consistent with a modified on-the-job search model where workers have idiosyncratic non-pecuniary shocks associated with holding a particular job. A negative shock means that a worker will accept a job with a lower premium than his current job.

The offer function is

$$\psi_i^* = Z_i\Pi + \gamma\bar{\psi}_{b(i)0} + \zeta_{G(b(i))} + \eta_i.$$

The worker's mobility decision depends on the difference between the new offer

and the net utility associated with receiving premium their current premium,  $\psi_{i0}^*$

$$v_i^* = \psi_i^* - \psi_{i0}^*,$$

where

$$\psi_{i0}^* = \psi_{i0} - \phi_i.$$

An indicator for whether the move occurs is therefore

$$I_i = \mathbb{1}(v_i > 0).$$

In which case, the conditional expectation of the observed wage premium distribution is:

$$E(\psi_i | \psi_{i0}, Z_i, i) = Z_i \Pi + \gamma \bar{\psi}_{b(i)0} + \zeta_{G(b(i))} + E(\eta | I_i = 1, \psi_{i0}, Z_i, i).$$

So unbiased estimation of the offer function requires a correction for selection by  $\psi_0$ :

$$\begin{aligned} E(\eta | I_i = 1, \psi_{i0}, Z_i, i) &= E(\eta | \psi_i^* + \phi_i > \psi_{i0}) \\ &= E\left(\eta | \eta + \phi_i > \psi_{i0} - Z_i \Pi - \gamma \bar{\psi}_{b(i)0} - \zeta_{G(b(i))}\right). \end{aligned}$$

In the empirical work, I estimate the selection correction model above as well as models that simply control for  $\psi_{i0}$  through a linear term:

$$E(\psi_i | \psi_{i0}, Z_i, i) = Z_i \Pi + \gamma \bar{\psi}_{b(i)0} + \zeta_{G(b(i))} + \beta \psi_{i0}.$$

I find that the selection correction procedure has a very minor, statistically insignificant effect on the estimate of  $\gamma$ . I use variants of the simpler model in the quantile regressions.

## 2.4 Data

I analyze the model on work histories drawn from the Longitudinal Employer Household Dynamics (LEHD) Program of the U.S. Census Bureau. The data that allow me to identify local social interactions are administrative data on the worker's Census block of residence from the Statistical Administrative Records System (StARS). The outcome of interest is the wage premium paid by the firm to which a worker moves, and the proposed network quality measures are moments of the distribution of wage premia held by other workers in an individual's neighborhood. Data on wage premia come from estimates of the Abowd-Kramarz-Margolis (AKM) log earnings decomposition (Abowd et al. 1999). The final analysis sample includes workers aged 18-70 who resided in one of 30 large Metropolitan Statistical Areas (MSAs) during 2002-2003 with information on the wage premia for any job they held in that two year period. A complete list of MSAs included in this paper is included in Table 2.1.

Table 2.1: List of Metropolitan Statistical Areas Used

MSA	Pop. 2000	Pop. Rank 2003	Pop. Growth 2000-08
Austin-Round Rock, TX	1, 249, 763	38	0.322
Baltimore-Towson, MD	2, 552, 994	19	0.045
Charlotte-Gastonia-Concord, NC-SC	1, 330, 448	37	0.279
Chicago-Naperville-Joliet, IL-IN-WI	9, 098, 316	3	0.052
Dallas-Fort Worth-Arlington, TX	5, 161, 544	5	0.221
Houston-Sugar Land-Baytown, TX	4, 715, 407	8	0.215
Indianapolis-Carmel, IN	1, 525, 104	34	0.125
Jacksonville, FL	1, 122, 750	42	0.170
Kansas City, MO-KS	1, 836, 038	27	0.090
Los Angeles-Long Beach-Santa Ana, CA	12, 365, 627	2	0.041
Louisville-Jefferson County, KY-IN	1, 161, 975	43	0.071
<i>continued on next page</i>			

Table 2.1 (Continued)

MSA	Pop. 2000	Pop. Rank 2003	Pop. Growth 2000-08
Miami-Fort Lauderdale-Miami Beach, FL	5,007,564	6	0.081
Milwaukee-Waukesha-West Allis, WI	1,500,741	36	0.032
Minneapolis-St. Paul-Bloomington, MN-WI	2,968,806	16	0.088
Oklahoma City, OK	1,095,421	47	0.101
Orlando-Kissimmee, FL	1,644,561	29	0.249
Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	5,687,147	4	0.027
Pittsburgh, PA	2,431,087	21	-0.032
Portland-Vancouver-Beaverton, OR-WA	1,927,881	25	0.145
Richmond, VA	1,096,957	46	0.117
Riverside-San Bernardino-Ontario, CA	3,254,821	13	0.265
Sacramento-Arden-Arcade-Roseville, CA	1,796,857	26	0.174
San Antonio, TX	1,711,703	28	0.187
San Diego-Carlsbad-San Marcos, CA	2,813,833	17	0.067
San Francisco-Oakland-Fremont, CA	4,123,740	12	0.037
San Jose-Sunnyvale-Santa Clara, CA	1,735,819	30	0.048
Seattle-Tacoma-Bellevue, WA	3,043,878	15	0.099
St. Louis, MO-IL	2,698,687	18	0.044
Tampa-St. Petersburg-Clearwater, FL	2,395,997	20	0.141
Virginia Beach-Norfolk-Newport News, VA-NC	1,576,370	32	0.052

The LEHD data are built around the longitudinal employer-employee links represented by state Unemployment Insurance (UI) wage records which constitute the job frame. UI records cover approximately 98% of wage and salary payments in private sector non-farm jobs. The LEHD infrastructure makes use of the unique individual and employer identifiers from this system to track workers over time as they move from job to job, and to identify which workers share an employer. These data are augmented with demographic characteristics through administrative record and statistical links as well as to employer characteristics, including employer size, industry, and ownership type. For a complete description of these data, see Abowd et al. (2009).

Data on place of residence come from the StARS database. StARS is a Census Bureau program originally designed to improve intercensal population estimates as well as refresh its household sampling frame. As part of this mission, it incorporates administrative data from the IRS, HUD, Medicare, Indian Health Service and the Selective Service to update information on residential geography once a year. Once geocoded, these data provide information on place of residence down to Census Block. Geocodes of this precision are available for at least 90% of all LEHD workers who appear in one of the 30 sample MSAs during 2002-2003.

To measure employer specific wage premia, I use estimates from a decomposition of log earnings into components associated with individual and employer heterogeneity. This decomposition as applied to matched employer-employee data was first introduced by Abowd et al. (1999) as a means of correcting biases in the estimation of industry and other more aggregated types of wage premia. The Abowd-Kramarz-Margolis (AKM) decomposition is an estimate of the model

$$\ln Y = X\beta + D\theta + F\psi + \varepsilon. \quad (2.19)$$

Here, this model is estimated on the set of all LEHD work histories for workers aged 18-70.  $Y$  is a vector of annualized earnings on the dominant job, and  $\varepsilon$  is a statistical residual.  $D$  and  $F$  are design matrices of the worker and employer effects.  $X$  is a matrix of time-varying controls consisting of a quartic in experience, year effects, and the exact within-year pattern of positive earnings. All of these measures are interacted with sex. The estimates used in this paper were conducted as part of the Human Capital Estimates Project within LEHD according to the estimation procedure described in Abowd et al. (2002; 2003). This

model is estimated on observations from 30 states between 1990-2003, covering 660 million wage records for 190 million workers and 10 million employers.

The job history information for workers includes information on transitions between dominant jobs. A dominant job in a given year is the one on which the worker had the most earnings in that year. For each dominant job, I merge the estimated employer-specific wage premium,  $\psi$ . Then for all workers who experience a job transition between 2002-2003, I identify the precise quarter of that transition. A worker can have one of three possible types of transition: out-of-sample to employment, employment to out-of-sample, and employer-to-employer. This paper focuses on employer-to-employer transitions. For out-of-sample to employment transitions, the date is the quarter in which the worker first has earnings with the new dominant employer. For employment to out-of-sample, the transition quarter is the last quarter with positive earnings reported from the dominant employer. Dominant job to dominant job transitions may occur annually given the definition of dominant employment. I refine the date of transition by finding the first quarter in which earnings with the new dominant employer exceed earnings with the old dominant employer.

The network quality measures are moments of the wage premium distribution for workers in one's residential neighborhood. According to the contagion model, the offer distribution depends on the distribution of jobs currently in progress among one's neighbors. To capture this, I measure network quality at the beginning of the quarter in which one's job transition occurs. The network quality measures are the means and variances of  $\psi_j$  for all workers residing in the relevant block, block group, or tract and who were employed the full quarter.



Table 2.2: Descriptive Statistics

Variable	Full Sample	Job Changers	Job Changers*
White	0.6572	0.6280	0.6220
Black	0.1151	0.1220	0.1205
Hispanic Origin	0.1167	0.1369	0.1400
Male	0.5098	0.4985	0.4979
Born in U.S.	0.8098	0.8098	0.8026
Age in 2002	40.5456	35.05848	34.9561
same_county	0.9765		
same_tract	0.8731		
same_block_group	0.8670		
same_block	0.8595		
Any transition	0.3116		
Transition to new job	0.0351	1	1
Transition out of sample	0.2634		
N	25,689,739	899,147	816,138

Table 2.2 presents descriptive statistics for the full sample as well as for the subsample of workers involved in a job-to-job transition. An observation in the sample is a worker from the LEHD infrastructure with positive earnings in at least one quarter of 2002-2003 who could be matched to a consistent block of residence in 2002-2003. For the urban workers that are the focus of the paper, this selection rule has little effect: over 95% of workers have consistent data on block of residence in both years. I require the recorded block of residence be in the same MSA in both years; that is, this analysis is for the group of workers who

do not move between MSAs during the sample period. I have recorded whether the worker experienced any kind of job transition. The second column in table 2.2 presents statistics for the subsample of workers who experienced a transition from one employer to another employer without an observed intervening spell out of the sample. In this sample of urban workers, The demographic characteristics are consistent with other published sources on labor force demographics. The sample of movers is marginally less white, more hispanic, and substantially younger, all of which are expected.

Thirty-one percent of workers experience any kind of job transition during the sample. For each worker meeting the inclusion criteria, I record the first observed job transition. Just 3.5% of workers experience a transition between dominant employers that involves no gap in the earnings history. This is significantly lower than would be expected from the reported rate of job-to-job transitions in other sources (Bjelland et al. 2008). However, this study considers transitions between dominant employers, where dominance is defined in terms of annual earnings. When a worker moves from one long-term employer to the next, it is picked up by this definition. However, many cases where a worker holds a short-term job between dominant employers will not be picked up using my definition. Bjelland et al. (2008), using a different definition of ‘main job’, find that roughly 31% of all transitions from jobs with tenure greater than one year are to jobs that last only 2-3 quarters. This could mean that as many as 12% of workers who appear to make a transition out of sample are actually transitioning into temporary jobs. This means that my sample of job-to-job transitions is more properly interpreted as a sample of immediate transitions from one relatively long-term dominant job to another. Given the objective of the study, this is the correct set of transitions to focus on. In transitions from one long-term job

to the next, it is more likely that the on-the-job search story will be accurate. A worker who takes a stop-gap job in between long term employers is perhaps more likely to have separated from the previous employer for other reasons or adopting a different kind of search strategy.

Table 2.2 also provides evidence in support of the identifying assumption that residential sorting operates at a higher level of geography than the individual block. I observe workers who move within the city over the two year period. 98% of workers in this sample remain in the same county, 87% remain in the same tract, and 86% remain in the same block. These categories are not exclusive, so the 13% of tract movers are also counted as block movers. Hence, the vast majority of workers who move, move to a new block group or tract, indicating that mobility decisions are not made on the basis of block-level characteristics.

## 2.5 Stylized Facts

I first provide evidence that the estimated distribution of  $\psi$  conforms to two major predictions of the job search model with contagious job information. In this section, I show that workers are more likely to move to jobs better than the one they currently have. Furthermore, the distributions of wage premia on destination jobs conditional on the premium in the origin job are strictly ranked in the sense of stochastic dominance. This is the first evidence that there is any mobility-related structure to  $\psi$  when estimated from the AKM decomposition.

I next show evidence of spatial correlation in the wage-premia held by workers. I compute non-parametric estimates of the spatial autocorrelation function

for block- and tract-level averages of log earnings and the components from the AKM decomposition. These reveal positive spatial autocorrelation in estimated wage premia.

These results should allay concern about a potential objection to the model. The empirical method is a two-stage estimation procedure. The first stage consists of estimating the empirical wage premia,  $\psi$ , from the AKM decomposition, and the second stage the estimation of the realized offer distribution from data on workers making direct dominant job transitions. This procedure is consistent under the assumption that the errors in the earnings equation are not correlated with errors in the job mobility process. This exogenous mobility assumption is a feature of the extended on-the-job search model developed in this paper. It is nevertheless a strict one. That the estimated wage premia conform to the stylized predictions of the model means that assumption may not be too strong.

### **2.5.1 Evidence of a job ladder**

Table 2.3 shows the fraction of workers that move to a job at the same decile, or a higher decile of the empirical  $\psi$  distribution than their current job. This probability is always strictly above 0.58, and significantly higher for workers starting from jobs with log wage premia in the lowest deciles. This evidence is consistent with the job ladder prediction of the basic search model developed above. Additional details on the nature of the job ladder evidence in these data appear in the corresponding figures 2.1 and 2.2. Figure 2.1 plots the cumulative frequency of destination wage premia for all job transitions stratified by decile of the origin job wage premium. The plots show decile-to-decile transitions, but

Table 2.3: Unconditional Transition Probabilities

Origin $\psi$ -decile $\psi_0^d$	$Pr(\psi_1^d \geq \psi_0^d)$
1	1
2	0.94
3	0.85
4	0.79
5	0.73
6	0.73
7	0.69
8	0.61
9	0.58
10	0.59

the same results hold when looking at more detailed quantiles. First, note that there is a clear first-order stochastic dominance relationship among the conditional distributions. Workers starting from jobs with higher wage premia are more likely to move to jobs with better premia. Second, for each conditional distribution, the probability of moving to a job with the same or a higher premium is always strictly higher than the probability of moving to a job with a lower premium.

Figure 2.2 plots the transition matrix between deciles of the wage premium distribution. Each ribbon shows, for workers whose initial wage premium fell in a certain decile, the fraction of transitions to jobs in each decile of the wage premium distribution. The saddle shape indicates that workers tend to move to

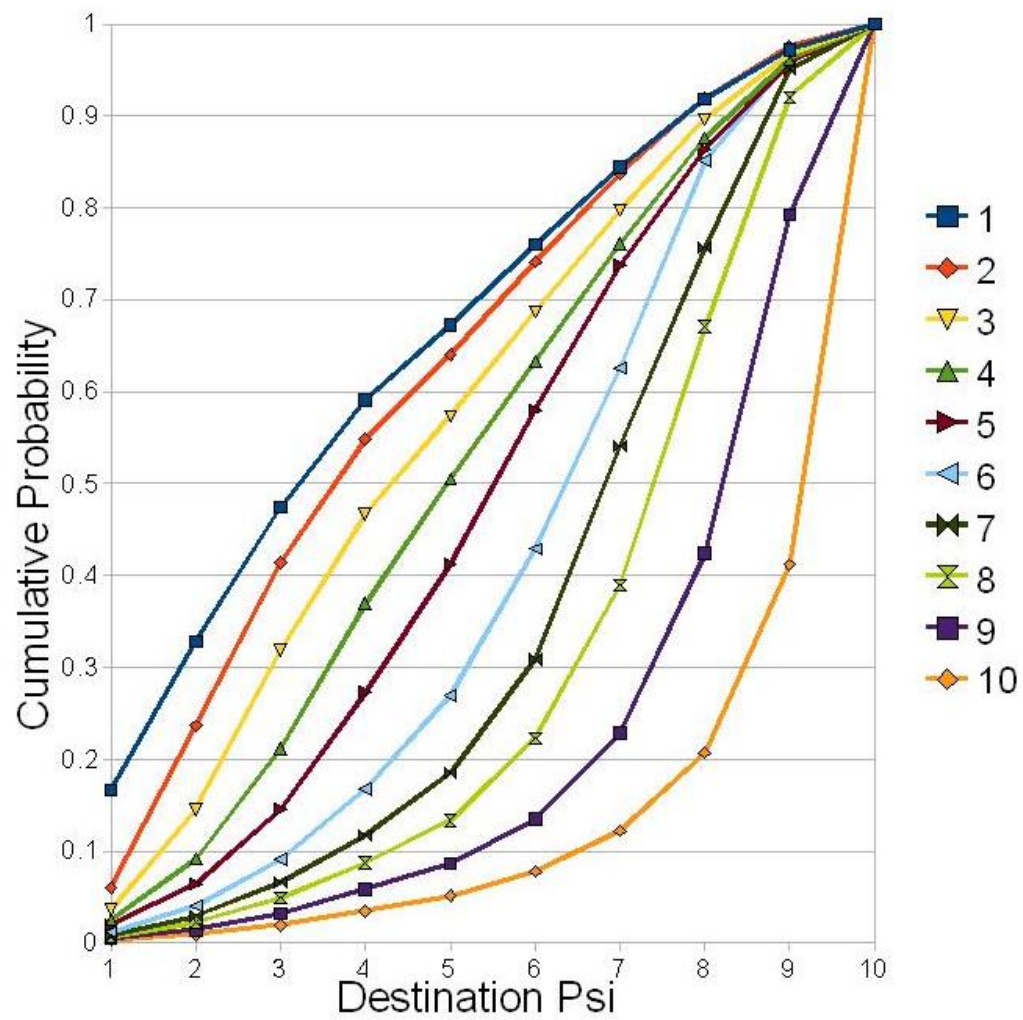


Figure 2.1: Cumulative probability of transition to each decile of the wage premium ( $\psi$ ) distribution, by decile of origin

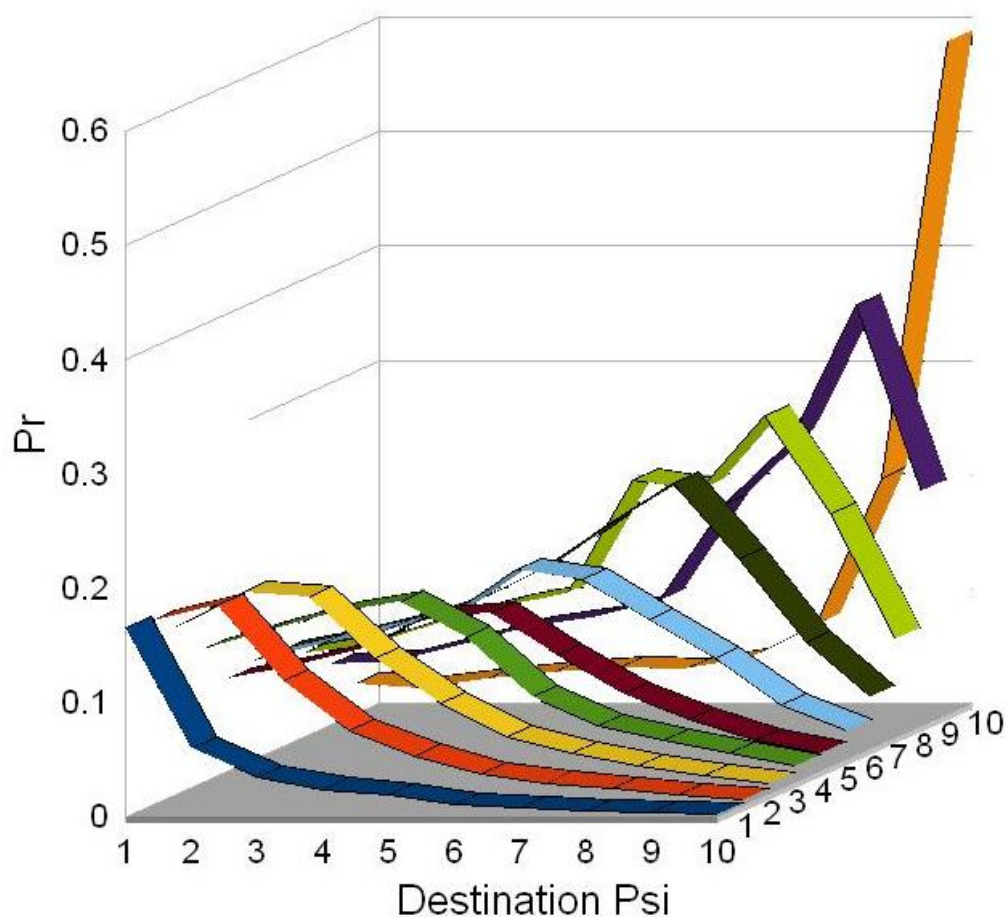


Figure 2.2: Probability of transition to each decile of the wage premium ( $\psi$ ) distribution, by decile of origin

jobs similar to, or better than the jobs they already have. The conditional densities are all peaked at the origin decile. This suggests that a mass of transitions are to jobs in the same decile. This is not a feature of the basic on-the-job search model. It instead suggests a relationship between the current job and the offer distribution. This will be the case when there is worker-level heterogeneity of any kind in the offer distribution. If, for instance, native workers tend to find

jobs with better wage premia, this will show up in the unconditional densities plotted here as correlation between current and future offers.

### **2.5.2 Evidence of spatial correlation in $\psi$**

The data must also exhibit correlation in wage premia between workers along the dimensions of social interaction that I can measure. If there is no spatial correlation in  $\psi$  at the Census block level, then the model can not be valid, either because social interactions do not function in the manner assumed in the model, or because social interactions are not strong at the block level. To evaluate this implication of the model, I compute the spatial autocorrelation function for each of the components of earnings from the AKM decomposition, both as tract-level and block-level means. In addition, I estimate the amount of clustering that appears in the data at the tract- and block-level as the amount of variation in the data that can be explained with tract- or block-level effects. Both measures provide evidence of spatial correlation in  $\psi$  for close neighbors. To my knowledge, these are the first estimates of their kind using matched employer-employee data for the U.S. Furthermore, the spatial autocorrelation estimates are the first of their kind to be estimated on earnings data at high spatial resolution.

#### **Estimates of the spatial autocorrelation function**

Figures 2.3 and 2.4 plot averages of the estimated spatial autocorrelation function in each MSA for tract- and block-level means of log earnings, the estimated person effect  $\theta$ , the estimated wage premium  $\psi$ , and the residual from the AKM decomposition,  $\varepsilon$ . The discussion in this section closely follows Conley and



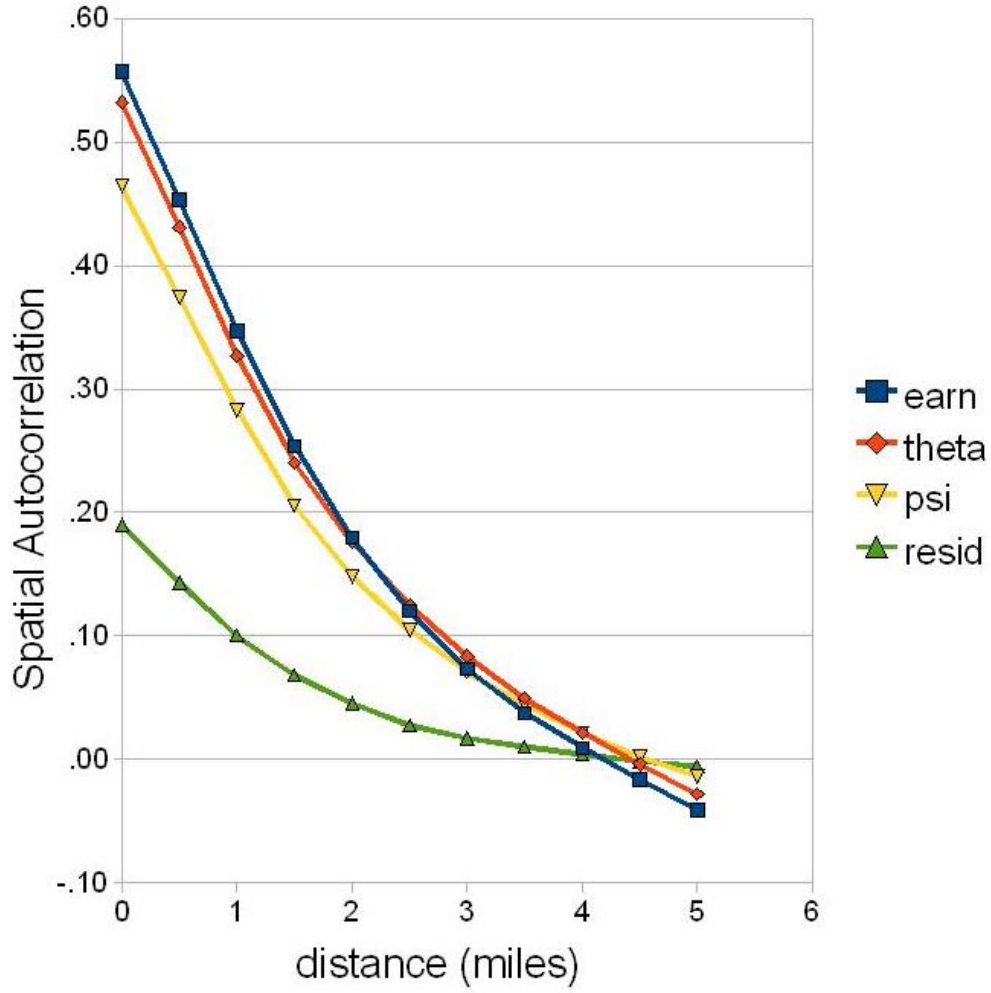


Figure 2.3: Spatial Autocorrelation Function: tract-level means

Topa (2002a) from which the method for this analysis was derived. The core statistical model is one in which random variable  $x_i$  is associated with a spatial coordinate,  $s_i$ . The spatial process generating the data is one in which the correlation between  $x_i$  and  $x_j$  depends only on the distance between  $s_i$  and  $s_j$ .

$$\text{corr}(x_i, x_j) = f(\|s_i - s_j\|) \quad (2.20)$$

This assumes that the correlations do not depend either on the precise loca-

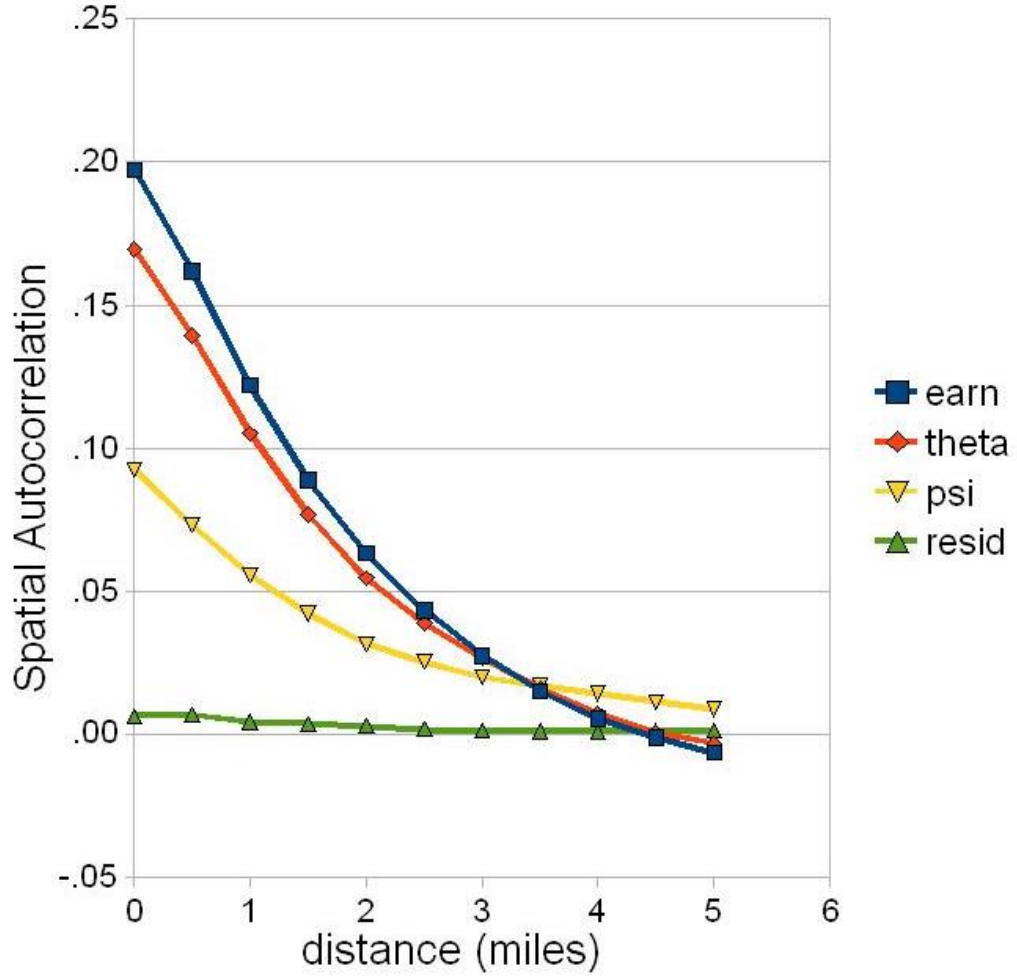


Figure 2.4: Spatial Autocorrelation Function: block-level means

tions in space of these random variables, nor the direction of the vector between them. I estimate the spatial autocovariance function at distance  $\delta$ ,  $f(\delta)$ , non-parametrically by

$$\hat{f}(\delta) = \sum_{i=1}^N \sum_{i'=1}^N \phi \left[ \frac{|\delta - A_{ii'}|}{\sigma} \right] (X_i - \bar{X})(X_{i'} - \bar{X}) \quad (2.21)$$

where  $A_{ii'}$  is the distance between  $i$  and  $i'$ .  $\phi()$  denotes the standard normal kernel. The spatial autocovariance function is estimated as the kernel-weighted

average of the products of demeaned observations. To convert this to the spatial autocorrelation, one must divide the resulting estimate by relevant product of standard deviations. With the normal kernel, this is just the sample variance.

I implement this estimator for tract-level and block-level means of all earnings and the components of the AKM decomposition. I compute  $\hat{f}(\delta)$  at distances from 0 to 5 miles at half-mile gridpoints.  $A_{ii'}$  is measured as the great-circle distance between internal points of the block or tract. For the block level estimates, the bandwidth parameter,  $\sigma$ , is set to 0.5. For the tract level estimates, it is set at 0.7. Since the computation scales in the square of the number of observations, for the block level calculation some simplification is required. I randomly sample block pairs at the rate of 1/100. For a hypothetical MSA with 5,000 blocks, which would be a fairly small one for this study, this means the spatial autocorrelation function is estimated from approximately 125,000 unique data points. To satisfy the disclosure avoidance restrictions required to publish these results, each point in the figures represents the unweighted average of the estimated  $\hat{f}(\delta)$  across 30 MSAs. There is some variation between the MSA-level estimates, but not enough to change the qualitative features of the plot. These plots are representative of most of the individual MSAs.

Both figures clearly show positive spatial autocorrelation in the tract- and block-level means of earnings,  $\theta$ , and  $\psi$ . The main point to take away is that spatial correlation in the estimated wage premia exist of workers in nearby blocks. This is consistent with the social interactions model of this paper. To be clear, there are also many other models that could generate these correlation patterns. The key challenge given the stylized fact is to identify the effect of social interactions in wage premia separately from other spatially correlated influences that

could produce the result.

These results contain a wealth of interesting information beyond the analysis in this paper. I mention just two points briefly. First, the block-level estimates show no spatial autocorrelation in the block level average residual. This is consistent with the model of the paper, in that it suggests that whatever process puts people in a particular block is not correlated with the earnings residual. Second, these plots give evidence on the relationship between the spatial correlation in earnings and sorting on unobservables. The spatial correlation in earnings is mirrored almost exactly by the spatial correlation in estimated person effect, which captures the effect on earnings of unobserved and observed non-time-varying characteristics. These results confirm the findings of Combes, Duranton and Gobillon (2008) that sorting on observable and unobservable human capital characteristics explain a large amount of the spatial wage distribution in cities.

### **Estimates of block-level clustering**

Neighborhood effect regression provides an alternative way to measure the extent to which workers are clustered with respect to the components of earnings. I present the results for tract- and block-level sorting on earnings, and the components of earnings estimated from the AKM decomposition. These results are an informative complement to the spatial autocorrelation results in the previous section. The results here estimate the amount of clustering within blocks, whereas the previous estimates showed the amount of spatial correlation between block-level averages. Following Ioannides (2004), Let  $k$  index neighborhoods, defined potentially as blocks, block groups, or tracts. Consider the  $R^2$

Table 2.4: Sorting by log earnings and components from the AKM decomposition,  $R^2$  method.

		Log Earnings	$\theta$	$\psi$	$\varepsilon$
tract	Mean	0.168	0.100	0.036	0.009
	Stan. dev.	0.0374	0.0244	0.0170	0.0030
block	Mean	0.291	0.203	0.119	0.082
	Stan. dev.	0.0357	0.0271	0.0171	0.0127
	Num. of MSAs	30			
	Num. of Obs.	14, 855, 153			

from the linear model

$$y_i = \eta_{K(i)} + e_i, \quad (2.22)$$

where  $e$  is an *i.i.d.* error measuring the dispersion in  $y_i$  around its neighborhood-level mean. The  $R^2$  measure is intuitively a measure of neighborhood clustering or sorting on  $y$  (Kremer and Maskin 1996; Kremer 1997). If there is no clustering by neighborhood, then the distribution of earnings within neighborhoods is just an exact copy of the city-wide distribution and the neighborhood effects shouldn't explain anything. If there is perfect sorting, so that each neighborhood distribution is point mass at one earnings level, then the neighborhood effects will explain everything.

For this paper, the important question is whether there is clustering within blocks in terms of the estimated wage premia,  $\psi$ . Table 2.4 presents evidence that this is the case. Each entry in the table is the simple average across the 30 MSAs

of  $R^2$  from estimating the model in equation 2.22 for a particular combination of geographic resolution (block or tract) and variable of interest (earnings,  $\theta, \psi$  or  $\varepsilon$ ). On average, block-level effects explain about 12% of the variation in wage premia. Tract effects explain 4%.

(Davidoff 2005) shows that these estimates are biased downward if earnings are measured with error, but are potentially biased upward by the mechanical relationship between  $R^2$  and the number of variables in the model. These models are estimated on very large samples, so adjustments of  $R^2$  for the number of blocks or tracts do not change the results. Measurement error is a potential problem, but the administrative earnings data used in this paper are generally more accurate than the self-reports of earnings in survey data.

The results for earnings are of independent interest. The block specification explains 29% of variation in measured earnings in the average MSA. The tract specification explains 18%. These results confirm the overall low levels of residential sorting with respect to earnings, or conversely, the high levels of mixing. Interestingly, the results presented here for blocks are very close to the results presented by Ioannides (2004) for tract-level estimates of sorting by household income from the American Housing Survey Metropolitan Sample. The most likely source of the discrepancy is the difference in variables being measured. Families are more likely to sort by household than individual income.

## 2.6 Estimation Results

Having established that the estimated wage premia,  $\psi$ , are consistent with the broad predictions of the model, I now use these data to estimate the influence

of local referral network quality on job search outcomes. Table 2.5 collects the results from estimating linear and quantile regression specifications of the form

$$\psi_i = Z_i\Pi + \beta\psi_{0i} + \gamma\bar{\psi}_{b(i)0} + \varphi\bar{\psi}_{G(b(i))} + \zeta_{G(b(i))} + v_i \quad (2.23)$$

Primary interest lies with estimates of the parameter  $\gamma$ , which measures the effect of local interactions on job offers.  $Z_i$  is a vector of individual characteristics including age and its square, indicators for whether the worker is white or not, hispanic or not, and male or not, as well as the estimated person effect,  $\theta$  from the first-stage. The notation  $b(i)$  indicates the Census block in which  $i$  resides.  $\bar{\psi}_{b(i)0}$  is the within-block average wage premium across all employed workers whose jobs were already in progress before the quarter in which  $i$  makes a transition, and that remained in progress in the quarter after.  $\psi_{0i}$  is the wage premium of the employer from which  $i$  transitions.  $\zeta_{G(b(i))}$  is a reference group effect where the notation  $G(b(i))$  indicates the reference group of contiguous blocks containing  $b(i)$ . The reference group in these estimates is the Census block group.

The key result is in the contrast between the baseline specification, which does not control for reference group correlations in outcomes, and the two specifications that do. Inference in the conditional mean regressions is based on heteroscedasticity-corrected standard errors that have been clustered at the MSA level<sup>1</sup>. The baseline model presented in the first column of table 2.5 shows the raw correlation between  $\bar{\psi}_{b(i)0}$  and  $\psi_i$ , the premium on the job to which  $i$  makes a transition, controlling for the premium on the origin job and observable characteristics that may influence formal search. The point estimate on  $\gamma$  in the base-

---

<sup>1</sup>This specification is very conservative. Under the empirical model, clustering at the county or tract level would be appropriate. As Cameron et al. (2008) point out, asymptotic tests based on data with around 30 or fewer clusters may over-reject. Even with standard errors cluster on 30 MSAs, the point estimates of interest are significantly different from zero in all cases. Alternative specifications that cluster on county or tract, which are available upon request, do not alter the qualitative results.

Table 2.5: Regression Estimates

Premium on next job, $\psi$	Baseline	Block Group Means	Block Group & Tract Means	Block Group Controls
Initial premium: $\psi_0(\beta)$	0.46* (.008)	0.45* (.008)	0.45* (.008)	0.45* (.008)
Mean premium in block: $\bar{\psi}_{block}(\gamma)$	0.33* (.016)	0.10* (.011)	0.10* (.011)	0.10* (.012)
Mean premium in block group: $\bar{\psi}_{bg}(\phi)$		0.34* (.024)	0.20* (.017)	
Mean premium in tract: $\bar{\psi}_{tract}$		0.15* (.025)		
white	0.001 (.001)	0.001 (.002)	0.002 (.002)	-0.002 (.001)
Hispanic Origin	-0.02* (0.005)	-0.01 (0.005)	-0.014* (.005)	-0.017* (.004)
male	0.03* (0.004)	0.03* (0.004)	0.04* (.004)	0.04* (0.004)
age in 2002	0.01* (.001)	0.01* (.001)	0.01* (.001)	0.01* (0.001)
Square of age in 2002	-0.00* (.000)	-0.00* (.000)	-0.00* (.000)	-0.00* (.000)
Born in U.S.	0.00* (.003)	0.01 (.002)	0.01 (.002)	0.01* (.003)
Person-effect from wage eqn: $\theta$	-0.00 (.009)	-0.01 (.010)	-0.01 (.010)	-0.01 (.010)
block group controls	no	no	no	yes
$N$	815, 899	815, 889	815, 889	815, 899
$R^2$	0.3149	0.3175	0.3176	0.2711



line model of 0.33 is on the same order of magnitude as the point estimate of  $\beta$ . In this specification, though,  $\gamma$  is absorbing any unobserved correlates of formal job search that aren't included in the model.

The social interaction parameter,  $\gamma$ , is identified in the model with reference group controls presented in the fourth column of table 2.5. The point estimate is  $\hat{\gamma} = 0.10 \pm 0.01$ , and is statistically significant. To interpret the point estimate in terms of the model, this means that 10% of job offers arrive through referrals. This is in line with Ioannides and Loury's (2004) analysis of referral use by workers in the Panel Study of Income Dynamics, which shows that 8.5 percent of employed workers report using referrals to search for work.

The other columns in table 2.5 present alternative estimates of  $\gamma$  based on a contrast between  $\bar{\psi}_{b(i)0}$  and  $\bar{\psi}_{G(b(i))}$ . The point estimates are nearly identical, and I conclude that the coefficient on the group-level average log wage premium has absorbed all of the unobserved correlation in outcomes. Because of its computational simplicity, I use this contrast to estimate the selection correction model as well as the quantile regressions.

As a check for robustness of my estimate of the social interaction parameter  $\gamma$ , I estimate the full econometric model of job-to-job mobility described in section 2.3.1. It allows for sample selection driven by the fact that only workers who receive sufficiently attractive offers change jobs. The attractiveness of a job offer depends on the wage premium of one's current job,  $\psi_0$ . Following the theoretical model,  $\psi_0$  is excluded from the offer function, but does appear in the selection equation. I estimate the selection correction model using data on all employed workers at risk to change jobs in 2002 quarter 4. The results are presented in table 2.6. As expected, in the selection equation, the log wage premium on the

Table 2.6: Selection Correction Model Estimates

Premium on next job, $\psi$	Offer	Selection
Selection on job-to-job move	Function	Equation
Initial premium: $\psi_0 (\beta)$		$-0.58^*$ (.017)
Mean premium in block: $\bar{\psi}_{block} (\gamma)$	$0.11^*$ (.023)	$0.10^*$ (.020)
Mean premium in block group: $\bar{\psi}_{bg} (\phi)$	$0.64^*$ (.060)	$0.32^*$ (.069)
$\lambda$ (Inv. Mills)	$0.48^*$ (.058)	
$\rho$	$0.79$	
$\sigma$	$0.61$	
$N$	1, 330, 475	
$\chi^2_{(9)}$	683.23	

Bootstrapped standard errors clustered on 30 MSAs. \* entries have p-value < 0.025. Both models include all controls from Table 2.5.

worker's initial job  $\psi_0$  has a strong negative effect on the probability of a job-to-job move. Workers living on blocks with better than average network quality for their neighborhood are also more likely to make a job-to-job transition. The point estimate on the social interaction parameter in the selection correction model is  $\hat{\gamma} = 0.11 \pm 0.02$ .

One objection to my research design is that the local referral interactions I model are most relevant to certain kinds of jobs, and are more likely to be used

Table 2.7: Heterogeneous Referral Effects

Premium on next job, $\psi$		
Mean premium in block: $\bar{\psi}_{block}$	0.10*	0.17*
	(.012)	(.023)
Born in U.S. $\times \bar{\psi}_{block}$		-0.09*
		(.021)
block group controls	yes	yes
$N$	815,899	815,889
$R^2$	0.2711	0.2712

Standard errors are clustered on 30 MSAs. \* entries have p-value < 0.025. Models include controls from Table 2.5.

by certain groups of workers. Previous research indicates that the use and efficiency of referrals differ considerably by demographic group. Furthermore, the kinds of jobs that are shared among residential neighbors are more likely to be jobs with relatively little specific skill requirements. To check whether my results are sensitive to heterogeneity in the effect of referrals, I estimate the model with block group controls and allowing for the use of referrals to be different for native workers and non-native workers. The results, reported in table 2.7, show that non-native workers have  $\hat{\gamma} = 0.17 \pm 0.02$ , which is a 70% increase over the pooled estimate. This finding is consistent with other work finding that immigrants are more likely to find jobs by referral than their native counterparts.

Observable demographic characteristics explain relatively little of the variation in the data. The signs on the coefficients associated with demographic and human capital characteristics have the same sign as would be expected in a Min-

cerian wage regression, but with only marginal significance in most cases. All of these estimates are an order of magnitude smaller than the point estimates of the social interaction parameter  $\gamma$ , and the effect associated with the initial job type. This is likely because these human capital effects are implicitly controlled in the first stage when  $\theta$  and  $\psi$  are estimated. Evidence of and absence of correlation between the estimated wage premia  $\psi$  and  $\theta$  matches evidence documented in other studies of the AKM decomposition. This result is consistent with the arrival of information about wage premia being only weakly related to individual ability, which is in turn consistent with the notion that they are non-economic rents associated with information frictions in the labor market.

Table 2.8 presents estimates of conditional quantile specifications for the 10<sup>th</sup>, 25<sup>th</sup>, 50<sup>th</sup>, 75<sup>th</sup> and 90<sup>th</sup> percentiles of the  $\psi$  distribution for job changers. The quantile regression results confirm the theoretical predictions of propositions 3 and 4. The key result is the pattern in the coefficient estimates associated with  $\bar{\psi}_{b(i)0}$ ,  $\bar{\psi}_{G(b(i))0}$  and  $\psi_0$ .

Let  $\beta(q)$  be the coefficient associated with  $\psi_0$  in the conditional regression of the  $q$ th quantile, and define  $\gamma(q)$  similarly. Propositions 3 and 4 predict that for  $q < q'$ ,  $\beta(q) > \beta(q')$  and  $\gamma(q) < \gamma(q')$ . In the search model, increasing  $\psi_0$  affects outcomes by increasing the reservation offer that triggers mobility. Intuitively, increasing  $\psi_0$  will have a larger impact on the distribution of acceptable offers close to the truncation point than those further away. The estimates clearly show  $\beta(0.1) > \beta(0.5) > \beta(0.9)$ . However,  $\gamma(0.1) > \gamma(0.25) = \gamma(0.5) < \gamma(0.75) < \gamma(0.9)$ . The predicted pattern appears if one considers the estimates associated with the neighborhood (block-group) level mean,  $\bar{\psi}_{G(b(i))0}$ . It is also impossible to reject the hypothesis that  $\gamma(0.1) = \gamma(0.25) = \gamma(0.5)$ , so the data weakly support the

Table 2.8: Quantile Regression Estimates

Premium on next job, $\psi$	Quantile				
	$q(0.1)$	$q(0.25)$	$q(0.5)$	$q(0.75)$	$q(0.9)$
Initial premium: $\psi_0$ ( $\beta$ )	0.59* (.002)	0.59* (.001)	0.50* (.001)	0.39* (.001)	0.32* (.002)
Mean premium in block: $\bar{\psi}_{block}$ ( $\gamma$ )	0.09* (.010)	0.08* (.006)	0.08* (.005)	0.11* (.005)	0.13* (.007)
Mean premium in block group: $\bar{\psi}_{bg}$ ( $\phi$ )	0.17* (.012)	0.22* (.007)	0.30* (.006)	0.40 (.007)	0.57* (.010)
$N$	815, 899	815, 889	815, 889	815, 899	815, 899
pseudo- $R^2$	0.1908	0.2187	0.2002	0.2711	0.1741

\* entries have p-value < 0.025. Models include all controls from Table 2.5.

proposition.

## 2.7 Conclusion

I find strong evidence of local social interactions in the transmission of information about employer specific wage premia. Workers whose neighbors have jobs paying higher wage premia are more likely to experience a job transition, and when they do, are more likely to move to a job with a better premium. I estimate this effect using variation in network quality among workers who reside in the same Census block group. This is identified as a social interaction effect assuming that workers cannot sort by block within block group, and that factors that cause search outcomes to be spatially correlated are homogeneous within block groups. The best estimate from the model indicates that 10% of a worker's job offers come from referrals. This is consistent with figures reported by other authors on the extent of referral use. These are the first results on direct local interactions in earnings outcomes in the context of a job search model. They complement existing work on local interactions in employment status and hours of work. This is also an important extension of the identification strategy of Bayer et al. (2008), which uses a similar identification strategy to establish local interactions in job-finding.

To motivate and structure the empirical work, I construct a model of job search augmented to allow for transmission of job information through referral network contacts. In addition to the main result, I estimate mean and quantile regression models to test the model's predictions. I show that the distribution of wage premia across workers who make job-to-job moves responds to variation

in referral network quality in a manner consistent with this model.

The model also predicts that workers who switch jobs tend to move into jobs with higher wage premia than their current job, and that there will be correlation in the wage premia held by workers who are socially connected to each other. I show that the log wage premia estimated from matched employer-employee data exhibit both of these properties. This is the first evidence of mobility-related structure in employer wage premia estimated from matched employer-employee data. I also estimate the spatial correlation structure of earnings, employer-specific wage premia, and worker ability. The block-level analysis in this paper is among the most geographically detailed studies of sorting by earnings, human capital, and employer characteristics in U.S. cities and is relevant to those interested in residential sorting by earnings, human capital characteristics, and employer characteristics in urban labor markets.

My findings add to a growing body of evidence on the importance of social interactions for job search and labor market outcomes. The data support a model in which social interactions function to help exchange information about particularly attractive job opportunities among socially connected workers. This has implications for the distribution of earnings, and also for the efficiency of labor market matching. The details of these distributional and efficiency impacts are important areas for future research.

## CHAPTER 3

# FREE TO MOVE? A NETWORK ANALYTIC APPROACH FOR LEARNING THE LIMITS TO JOB MOBILITY

### 3.1 Introduction

There is a network of connections between employers and employees that evolves over time as people move from job to job. This realized mobility network connects workers in the economy with every employer with whom they have held a job. The advent of population-level matched employer-employee data means it is possible to document the realized mobility network empirically (Abowd et al. 2002). I argue that the structure of this network contains information of interest to labor economists. First, it provides a set of facts that should be respected in modeling job mobility. Second, it suggests a novel approach to identifying labor market clusters, whether they be sectoral, geographic, or discriminatory, in terms of partitioning the network where mobility is relatively weak. In this paper, I document several of the stylized facts of the realized mobility network in data from the Panel Study of Income Dynamics (PSID). In terms of degree distribution, clustering coefficient, edge-to-node ratios, and clustering, the realized mobility network is strikingly similar to other, unrelated networks that have been widely studied outside of economics. This is fortunate, because it means that it may be possible to use existing reduced form models of network formation to inform structural models of job mobility. I then go on to show that partitioning this network on the basis of structure alone into relatively tightly connected groups of nodes reveals labor market clusters that are differentiated along race, gender, and income. It follows that these tools may be



quite powerful for the unsupervised detection of labor market clusters.

Labor economics has ignored the relational structure of the realized mobility network, at least in empirical work, focusing instead on the role of agent attributes in determining labor market outcomes. Yet, for many basic questions, such as wage determination, this may be misguided. This paper goes to an opposite extreme by focusing completely on network structure and ignoring agent attributes. In so doing, it asks whether the network structure alone reveals economically salient facts about the labor market. A simple thought experiment reveals why this is so. Suppose we have to predict the wages of two identical workers about whom we know nothing of job history. The usual model might predict that each worker's wage is an independent draw from some wage distribution, conditional on human capital and demographic characteristics. Now, suppose we also know that these two workers ever worked for the same employer. Intuitively, this knowledge should also condition our forecast of the wage. Alternatively, what if we know they never worked for the same employer, but they both worked with people who themselves worked together? This 'short path' between work histories may be related to wage outcomes if, for instance, workers with similar histories tend to have similar earnings.

The analysis carried out here is for a sample of heads and spouses in the Panel Study of Income Dynamics (PSID) between 1987 and 1997. The set of 'pseudo-employers' from this data set is defined as the set of unique industry-occupation pairs that appear in the sample. The analysis of the PSID illustrates the range of techniques, and character of results that might be expected in matched data. Characterizing the topology of the realized mobility network in the PSID is the first contribution of this paper. Because the realized mobility

network is not a proper social network the particular statistics used to describe the network topology do not have clear economic interpretations. However, these statistics reveal information about the nature of the process through which the network was formed. For instance, it appears that the connections between workers based on work histories are more random than the connections between firms based on shared employment. The characterization of the network topology thus provides a handful of stylized facts that can be used in modelling labor market dynamics as a network formation process.

The second contribution of the paper is an application to the detection of labor market clusters of a method for detecting the ‘community structure’ of the realized mobility network. Mobility may be restricted between different sets of jobs for a variety of reasons. Theories of geographic immobility, occupation- or industry-specific human capital, and market segmentation all predict that observationally similar workers may have differing propensities to enter a particular job based on their prior employment history. In light of this, we would expect the realized mobility network to exhibit some more dense clustering amongst jobs that tend to be more closely related, with relatively fewer connections between those densely connected groups.

In the literature on complex networks, identifying these dense subgraphs is called a community structure problem. I first apply a method developed by Girvan and Newman (2002) to find community structure in the realized mobility network. The resulting partition of workers into different market clusters shows that mobility restrictions are correlated with demographic, gender and income differentials, which is consistent with existing models. I then develop and implement a second method for finding labor market segments based on

the principle of maximum likelihood. Using a simple Markov model of job mobility, I find the maximum likelihood partition of pseudo-employers. Like the first method, I find that this method delivers clusters that are correlated with observable and unobservable determinants of earnings. Altogether, these findings support the idea that there is interesting and exploitable labor market information in the structure of the realized mobility network.

## 3.2 The Realized Mobility Network

This paper works with a convenient representation of matched employer-employee data called the realized mobility network. The realized mobility network connects each worker with every employer she worked for during the sample. More formally, for each period of the sample we define the realized employment network as a bipartite graph,  $G_t$ , with vertex set  $V(G_t) = (\{1, \dots, I\}, \{1, \dots, m\}) = (I, M)$  and edge set  $E(G_t) = \{(i, m) \in I \times M | m = J(i, t)\}$ . The realized mobility network,  $G$ , is defined simply as the network whose vertex set is  $V(G) = (I, M)$  and whose edge set is  $E(G) = \cup_{t=1}^T E(G_t)$ .

### 3.2.1 Basic Graph Theory

A *graph* or *network*,  $G$  is defined by a given set of *nodes* or *vertices*,  $N(G) = \{1, \dots, n\}$  and a set of *edges*,  $E(G) \subset N \times N$ . When the edges are undirected we have  $(i, j) \in E(G)$  if and only if  $(j, i) \in E(G)$ . In this paper, all graphs are undirected, so the simplified notation  $ij \in E(G)$  will be used to say that  $i$  and  $j$  are connected in  $G$ . Note also that the terms *graph* and *network* can be used interchangeably; both

terms refer to the same theoretical construction.

Several alternative representations of  $G$  are possible. The adjacency matrix representation,  $A$ , is an  $n \times n$  matrix whose  $ij$ th entry is 1 if  $ij \in E(G)$  and 0 otherwise. Clearly, the adjacency matrix for an undirected graph is always symmetric.  $G$  can also be represented by the list of its edges, as long as each vertex in the graph is attached to at least one edge. Finally, the graph can be characterized by listing, for each vertex  $i$ , the *neighbors of  $i$*  in  $G$ . We let the set  $N_G(i) = \{j \in N : ij \in E\}$ . The degree of  $i$  is simply the number of its neighbors,  $|N_G(i)|$ .

Given  $i$  and  $j$ , we say that  $i$  and  $j$  are *path-connected* if there exists a set of edges in  $G$  connecting  $i$  and  $j$ .

Given a graph  $G$ , its components are the completely connected subsets of its node set. That is  $\hat{N} \subset N$  is a component of  $G$  if for each  $i$  and  $j$  in  $\hat{N}$ ,  $i$  and  $j$  are path connected.

The basic representation of the realized mobility network is as an undirected bipartite graph. A bipartite graph,  $B$ , is a graph with two non-intersecting sets of vertices with the feature that edges can only link nodes that are in different sets. In the context of the realized mobility network, one set of nodes represents the set of workers, and another set of nodes represents their employers. More formally, we define a bipartite graph,  $B$ , as a graph whose node set,  $N(B) = (N_1, N_2)$  where  $N_1$  and  $N_2$  form a partition of  $N$  and such that  $ij \in E(B) \rightarrow i \in N_1$  and  $j \in N_2$ .

In characterizing the realized mobility network, we also want to consider the set of connections it induces between firms, and between workers. We can

### Realized Employment Network, t=1

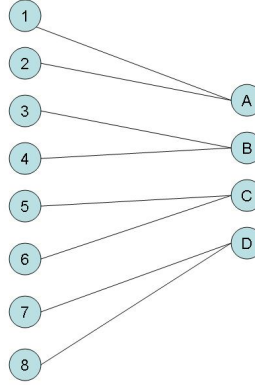


Figure 3.1: Realized Employment Network:  $t = 1$

think of two workers as being connected if they ever held a job with the same employer. Likewise, two employers are linked if they ever shared an employee. These induced unipartite graphs are called the *one-mode projections* of the realized mobility network. Given bipartite graph  $B$ , we may form the one-mode projection of  $B$  onto  $N_1$ ,  $B^{N_1}$ , as follows. Let  $N(B^{N_1}) = N_1$  and let  $vw \in E(B^{N_1})$  if and only if there exists  $j \in N_2$  such that  $vj \in E(G)$  and  $wj \in E(G)$ . The well-known game “Six Degrees of Kevin Bacon” is as clear an example as you could want of bipartite graph projection. The object of the game is to trace a path from a given actor, say Bela Lugosi, to Kevin Bacon where a connection exists between two actors if they were ever in the same film.<sup>1</sup> The underlying network is bipartite, connecting actors and films where the link means “x appeared in y”. The Kevin Bacon game is based on the projection of this graph onto the set of actors.

### Realized Employment Network, t=2

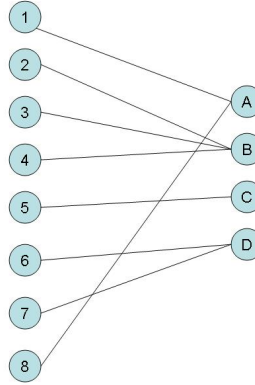


Figure 3.2: Realized Employment Network:  $t = 2$

### Realized Employment Network, t=3

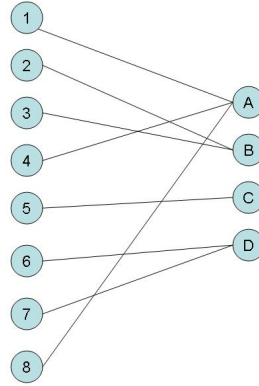


Figure 3.3: Realized Employment Network:  $t = 3$

## 3.2.2 Defining the Realized Mobility Network

The realized mobility network is generated from panel data of the form  $\{i, t, m\}$  where  $i \in W$  is the index of a particular worker,  $t \in T$  is the time period, and  $m \in M$  is an identifier of  $i$ 's employer at time  $t$ . The subsample at a fixed point in time,  $t$ , may be regarded as the set of edges in a bipartite graph,  $G_t$ .

---

<sup>1</sup>Bela Lugosi is three steps away from Kevin Bacon. See <http://oracleofbacon.org/> for details.

### Realized Mobility Network

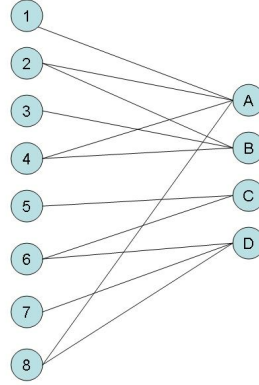


Figure 3.4: Realized Mobility Network

$G_t$  is called the *realized employment network* at time  $t$ . The vertex set of  $G_t$ ,  $V(G_t) = (W_t, M_t)$  where  $W_t \subset W$  and  $M_t \subset M$  are the sets of workers and employers observed in the particular period. An element  $ij$  of the edge set  $E(G_t)$  indicates that worker  $i$  holds a job with employer  $m$ . Given a sample period,  $t \in \{1, \dots, T\}$ , the realized mobility network,  $G$ , is simply the network formed by setting  $V(G) = \cup_{t=1}^T (W_t, M_t)$  and  $E(G) = \cup_{t=1}^T E(G_t)$ . Figures (3.1) through (3.4) illustrate the formation of a realized mobility network over three time periods in the simple setting where the set of workers,  $W = \{1, 2, 3, 4, 5, 6, 7, 8\}$ , and the set of employers  $J = \{A, B, C, D\}$ . This simple example shows that while the structure of  $G_t$  is rather sparse at any point in time,  $G$  becomes densely connected very quickly<sup>2</sup>.

<sup>2</sup>In general, these data are also matched to  $X_i$ , which are time-invariant characteristics of the worker such as race and education,  $X_m$ , the time-invariant characteristics of the employer,  $Z_{i,t}$ , the time-varying characteristics of the worker,  $Z_{m,t}$ , the time-varying characteristics of the employer, and  $Y_{i,m,t}$  which are job-specific characteristics that vary over time such as the wage and tenure.

### 3.3 Related Literature

Sociological analysis starts with the premise that people's behaviors are mediated by their social environment. Modeling the social environment as a network of relationships among otherwise independent people can be a fruitful way to explore the effects of social structure on individual and group behavior (Holland and Leinhardt 1977). Any market transaction forms a connection, even if only a fleeting one, between the people, firms, and other economic agents who are parties to it. But while the formalization of social relationships as networks has a long history in sociology<sup>3</sup>, the analysis of economic relationships as networks is virtually nonexistent. This may be for at least two reasons. First, although the relational properties of economic exchange are ubiquitous, they are generally speaking also unobserved in the data. In economic data, we are generally only ever informed about the specific characteristics of one agent in a transaction. Second, perhaps relatedly, most economic analysis abstracts away from the relational aspects of market interactions. These specific relational interactions are regarded as irrelevant in the determination of the market level phenomena of price, quantity and distribution of resources.

There are very few existing empirical analyses of network structures in economic data, and this is the first, to my knowledge, of labor market data. Within economics, most of the interest in social network analysis has been in developing theoretical models to explore how social network structures mediate the transmission and distribution of important economic information. Within labor economics, many papers analyze the effects of employee social networks on the transmission of information about various job opportunities (Montgomery

---

<sup>3</sup>Moreno (1934) is generally credited as the first network analysis of social relations. Wasserman and Faust (1994) provides a thorough synthesis and overview of social network methods.



1991; Calvo-Armengol and Jackson 2004). The empirical analyses associated with these theoretical models attempt to infer the impact of social network structure through identification of social interaction effects (Conley and Topa 2002b), but do not explicitly assess network structure, generally because data on social relationships are not available.

The analysis of the topology of a realized mobility network has more in common with the interdisciplinary literature on complex networks developed by computer scientists, physicists and sociologists. Within that literature, empirical analyses have been conducted on social networks, information, biological, and technological networks. None of these have looked at the structure of connections in labor market data, nor explored the possibilities of using these methods to detect structure in job mobility patterns. However, all of the topological properties that are considered in this paper are standard fare in the complex networks literature. The surveys written by Jackson (2005) and Newman (2003) provide excellent overviews of empirical studies of complex network topologies.

The problem of detecting community structure has a long history in social network analysis. It is related to the more general problem of finding partitions of a graph with various properties. The approach used in this paper was developed in a pair of papers by and Clauset et al. (2004). Within economics, Copic et al. (2009) have a recent paper analyzing community structure in a network formed from data on reference patterns in economics journals. Their approach is similar to the one used here in that it involves finding the partition that maximizes an objective function. In their case, the objective function is a likelihood function based on an underlying model of the graph formation process, and for

which they provide a decision-theoretic foundation. Interestingly, the modularity objective function is not equivalent in general to the likelihood function of Copic et al. (2009).

Community structure in the realized mobility network reveals patterns in job-to-job transitions suggesting that mobility is more likely between certain kind of jobs. This perspective is related to labor economics literature on the mobility of workers across industries and occupations, and to the literature on labor market segmentation. Kambourov and Manovskii (2008) find strong evidence that mobility within occupation is more likely than mobility between occupations, with similar but weaker results for inter-industry transitions. Both human capital theories and labor market segmentation theories predict differential transition rates between different kinds of jobs. The analysis of community structure in matched employer-employee data provides a method for detecting where these differential transition rates might lie, without making an *a priori* assumption on the nature of the barriers to mobility.

### 3.4 Data

My sample from the PSID includes all heads and their spouses from 1987-1997. An individual was in the sample as long as she was in a family that responded to the survey in both 1987 and in 1997. Individuals who never reported a primary industry-occupation pair during the sample are omitted, since they do not contribute a (relevant) vertex to the network. I chose heads and spouses because the PSID consistently collects their industry and occupation data. Additional demographic variables incorporated into the analysis below include race

(white or not), gender, labor income, family income, state of residence, age in 1994, and education. The main sample includes 9,224 individuals who report a valid industry-occupation pair for at least one year in the sample.

This main PSID sample is recorded as a person-level file which is easily converted to a bipartite adjacency list, and then to a bipartite edgelist. From the bipartite edgelist, I construct the one-mode projections  $G^W$  and  $G^M$ . The topological characteristics of all three networks are computed as discussed in the methodology section. To perform the community structure analysis using the FastCommunity algorithm of Clauset, Newman and Moore, it is also necessary to extract the components of each graph to be analyzed separately<sup>4</sup>. In the projection,  $G^W$ , there are 410,616 edges connecting the workers in  $W$ . Finally,  $G^W$  has one large connected component with 7,425 nodes and 381,599 edges.

For the mobility analysis, the main sample is further restricted to those individuals who were over 29 years of age in 1994 and who contribute at least two years of valid industry and occupation data. The restriction to persons over age 22 is to reduce the influence of early career mobility associated with career-shopping. The number of workers in this subsample is  $|W| = 7,515$  and the number of pseudo-employers (unique industry-occupation pairs) is  $|M| = 6,943$ . In analyzing job-to-job mobility, I focus on a decomposition of the strongly connected component (SCC) in the network over pseudo-employers. A strongly connected component is one in which it is possible to get from one job to any other job through a series of directed edges. The strongly connected component in the network consists of 5,103 pseudoemployers. 7,363 of the total 7,515 workers are ever employed in one of the pseudoemployers in the SCC. Table

---

<sup>4</sup>The SAS and C++ code for the network transformations and calculation of network statistics is available from the author. Code for the FastCommunity algorithm is available from the webpage of Aaron Clauset, <http://www.cs.unm.edu/~aaron/research/fastmodularity.htm>

Table 3.1: Data Summary

	PSID Sample	SCC*	CC**
white	0.690 (0.504)	0.690 (0.505)	0.690 (0.505)
male	0.512 (0.50)	0.509 (0.50)	0.509 (0.50)
number of jobs	4.5 (2.66)	4.5 (2.66)	4.5 (2.66)
average labor earnings†	\$19,900	\$19,759	\$19,860
age	38.83 (15.529)	38.81 (15.675)	35.43 (12.0951)
Less than HS	0.202 (0.4015)	0.202 (0.4015)	0.205 (0.4038)
High School	0.330 (0.4702)	0.327 (0.4693)	0.323 (0.4676)
Some College	0.231 (0.4215)	0.231 (0.4216)	0.234 (0.4232)
College	0.139 (0.3464)	0.141 (0.3479)	0.139 (0.3458)
Postgraduate	0.070 (0.2554)	0.071 (0.2569)	0.0952 (0.2935)
Number of Obs.	7,515	7,363	7,425

Standard errors are in parentheses.

\*Employed by a pseudoemployer in the strongly connected component of the realized mobility network.

\*\*The largest connected component of the realized mobility network.

† Median of average within-sample labor earnings.

3.1 shows that the basic characteristics of the two samples of workers are nearly identical. These workers initiate 31,578 unique job spells with different pseudoemployers for a total of 51,066 job-year observations in the SCC.

### 3.5 Basic Topology of the Realized Mobility Network in the PSID

Networks, social or otherwise, are complex when their size and patterns of attachment defy regularity or visual clarity. The development of large network datasets and the computational power to analyze them has driven a search for methods to characterize features of interest in complex networks. The three most commonly reported measures of network topology are the degree distribution, the clustering coefficient, and the average distance between nodes. In this paper, I focus on these three measures for consistency with the existing literature and explore whether these properties of the network topology have inherent economic interest. In general, they are difficult to interpret individually, but taken together tell us about the process of network formation.

#### 3.5.1 Degree Distribution

Recall that the degree of a node,  $i \in V(G)$ , which we denote  $k_i$ , is the number of neighbors  $i$  has in  $G$ . That is,  $k_i = |N_G(i)|$ . The degree distribution,  $F(k) = \Pr(k_i \geq k)$  is a fundamental concept linking the analysis of complex networks to the theory of random graphs. One of the recurrent observations of complex network research is that social networks have degree distributions that tend to follow a power law, and that tend to be scale free. This contrasts with basic models of random network formation, which result in a Poisson (when formation is static) or exponential degree distribution (when formation is dynamic). In this paper, I report whether the observed degree distributions are

closer to an exponential or a power law distribution based on the quality of the fit between the theoretical and the empirical distribution. I interpret a roughly exponential distribution as evidence of randomness in the statistical process by which the network formed. A roughly power law distribution suggests that there is some non-randomness in the sense that the probability of attachment is not independent of node degree.

In the realized mobility network, we can calculate the degree distributions of the worker nodes and the employer nodes separately. The degree distribution of the worker nodes is therefore a measurement of the distribution of the number of job changes observed during the sample. The degree distribution of the employer nodes is a measurement of the number of employment relationships in which a given employer was engaged during the sample. This gives the curious result that employers that are occupied by a large number of workers at any time and low turnover are topologically similar to employers with a low number of workers at any point in time, but high turnover.

In the one-mode projection  $G^W$ , the degree of  $i$  is the number of workers who worked for the same employer as  $i$  at any point in the sample, even if they did do so at the same time. As in the bipartite graph, the degree will be influenced by both the number and type of employers of each worker. A worker making few transitions but working for an employer with either high turnover, or high employment will have high degree. The distribution of degree in  $G^W$  nevertheless captures something about the way work experiences are shared across the population. The degree of a node  $m$  in  $G^M$  measures the number of other jobs that were held by workers who also worked for  $m$ . This measurement, too, confounds employer size and employer turnover, but nevertheless indicates

something about the generality or specificity of the particular job.

### 3.5.2 Clustering

Measures of clustering are meant to capture the extent to which connections in the network are transitive; that is, of the extent to which a given node's neighbors tend to also be neighbors of each other. In the social networking context, clustering in a graph captures the idea that 'friends of my friends are also my friends'. In slightly more abstract terms, substantial clustering in the graph should be reflected in a heightened number of 'triangles'. One statistic that has developed to capture the extent to which a large graph has this property is the 'clustering coefficient',  $C$ , which is a measure of the mean probability that the friend of your friend is also your friend. This has been measured both as

$$C = \frac{3 \times \# \text{ of triangles in graph}}{\# \text{ of connected triples}}$$

or

$$C_i = \frac{\# \text{ of triangles connected to } i}{\# \text{ of connected triples centered on } i}$$

and then

$$C = \frac{1}{n} \sum_i C_i$$

Where a 'connected triple' is a subgraph of three nodes,  $\{j, i, k\}$  in which  $j$  and  $k$  are both connected to  $i$ , which is called the 'center' of the triple. Only the latter measurement is used in this paper.

The concept of clustering does not apply to bipartite graphs, but it is possible to measure in each of the one-mode projections of the realized mobility network.

The clustering  $C_i$  of any node  $i$  in  $G^W$  is determined as follows. The degree of the node determines the number of triples centered on  $i$ . A triple  $\{j, i, k\}$  will be completed in one of two cases: if  $i, j, k$  all share the same employer, or if  $i$  and  $j$  share a different employer from  $i$  and  $k$ , but  $j$  and  $k$  also happen to share a third employer.

### 3.5.3 Distance

Finally, the geodesic distance between  $i$  and  $j$ ,  $d_{ij}$  is the number of edges along the shortest path between them. The mean geodesic is the sample average over all nodes.

$$l = \frac{1}{1/2n(n+1)} \sum_{i \geq j} d_{ij}$$

The mean geodesic in  $G^W$  is a measure of how closely linked two workers are in terms of their employment history. The length of the shortest path between  $i$  and  $j$  in  $G^W$  indicates the minimum number of transitions required to move from one of  $i$ 's employers to one of  $j$ 's. Similarly, the length of the shortest path in  $G^J$  between  $j$  and  $k$  tells us something about the ease of moving from  $j$  to  $k$ . Namely, if  $j$  and  $k$  have distance one, then someone was able to move from  $j$  to  $k$ . If they have distance two, then no one moved from  $j$  to  $k$ , but there is some intermediate job that connects the two. In some sense, the graph distance is a measurement of the ease with which a hypothetical worker could hope to make a transition between two jobs. This is connected to the idea of finding mobility segments, which I turn to in the next section.



### 3.5.4 Results

Statistics summarizing the topology of the bipartite realized mobility network,  $G$ , and its one-mode projections,  $G^W$  and  $G^M$  appear in Table 3.2. Statistics on average degree are reported for both sets of nodes in  $G$ . For  $G^W$  and  $G^M$  I also report the clustering coefficient,  $C$ , calculated as discussed above. To further characterize the degree distribution, I also report parameters for either the exponential distribution or power law distribution with the best fit. For comparison, I have also included statistics on two social networks with similar topological properties: the network of film stars and the network of coauthorship relationships in physics Newman (2003)

#### Degree distribution of $G$

In the bipartite realized mobility network,  $G$ , the average degree of the nodes in  $W$ ,  $\mu_W = 2.6$ , and the average degree of the nodes in  $M$ ,  $\mu_M = 4.3^5$ . A worker can only report one primary industry and occupation in each of the six years of the sample, so there is an upper bound of 6 on the degree of any node  $w \in W$ . There is no such limitation on the degree distribution of  $m \in M$ .

Graphs of the cumulative degree distributions for  $M$  and  $W$  appear as figures 3.5 and 3.6. The degree distribution of  $M$  is roughly linear for small degrees when plotted on log-log axes, suggesting a power law degree distribution in that portion of the support. However, the distribution is also heavily skewed

---

<sup>5</sup>Because each link with an end in  $W$  must also have an end in  $M$ , we necessarily have

$$\frac{\mu_W}{M} = \frac{\mu_M}{W}$$

Table 3.2: Basic Topology of the Realized Mobility Network in the PSID

Network		Nodes	Edges	Avg. Degree	Degree	Degree Dist'n	Clustering
Bipartite	Worker	9,224	22,469	2.6	2.6	Exp. (0.6)	n/a
	Job	5,255	22,469	4.3	4.3	Power (1.5)	n/a
Projections	Worker	9,224	428,848	102.22	102.22	Exp.(0.01)	0.71
	Job	5,255	26,614	9.5	9.5	Power (1.4)	0.57
Comparisons	Actors <sup>†</sup>	449,913	25,516,482	113.43	113.43	Power(2.3)	0.78
	Physics	52,909	245,300	9.27	9.27	n/a	0.56
	Coauthorship <sup>†</sup>						

<sup>†</sup> As reported in Newman (2003)

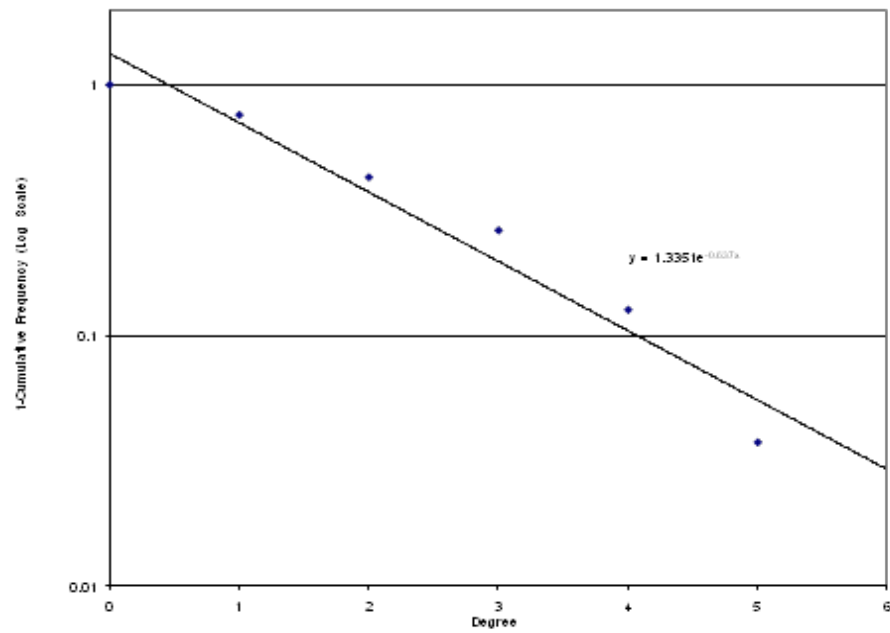


Figure 3.5: Degree distribution of the person nodes in the realized mobility network in the PSID

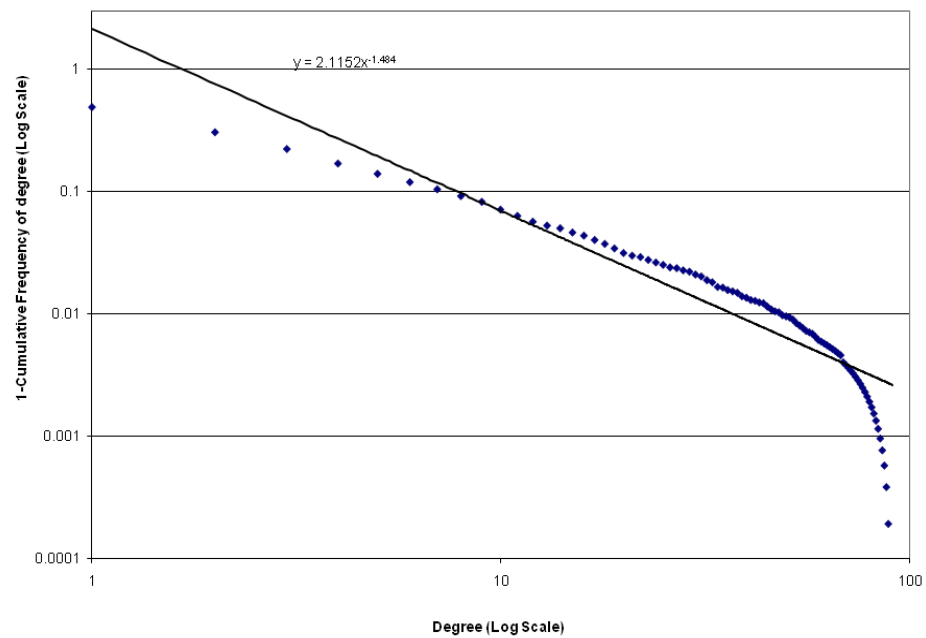


Figure 3.6: Degree distribution of the pseudoemployer nodes in the realized mobility network in the PSID

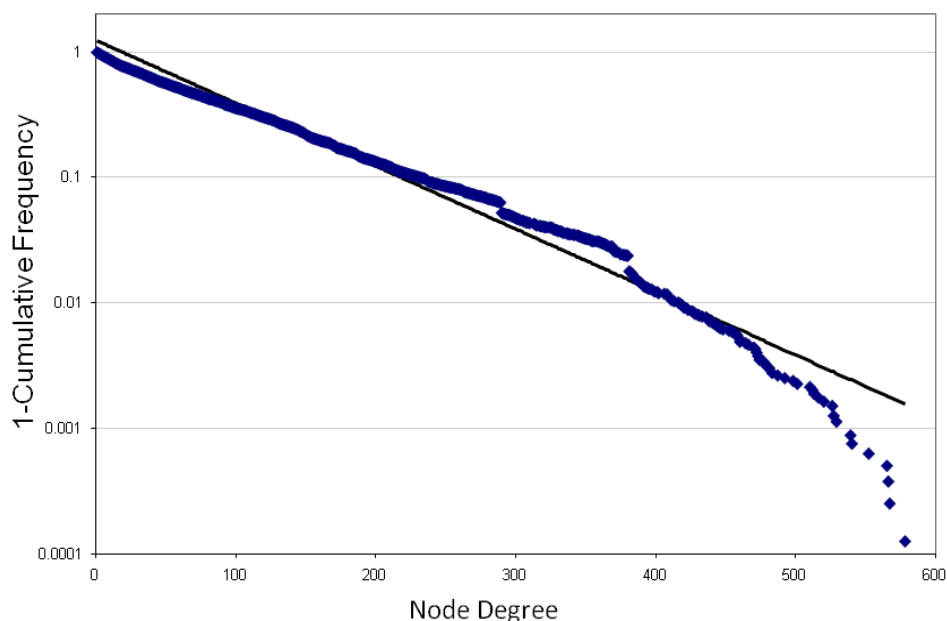


Figure 3.7: Degree distribution of the projection of the realized mobility network in the PSID sample onto workers

toward the right tail. This shape for a degree distribution is not uncommon in the literature (Newman 2003). The degree distribution in the set of worker nodes is better fit as an exponential. It is noteworthy that the functional form of the degree distributions match those of the corresponding one-mode projections.

### The Projection onto $W$ : $G^W$

Projection onto  $W$  produces a graph with a great deal of connectivity. This graph has 428,848 edges connecting the 9,224 worker nodes. Of these 9,224 nodes, 8,360 are included in one large connected component. The remainder of the components are singletons (corresponding to workers who had only one pseudo-employer that was never shared by any other worker) and dyads. The

average degree in  $G^W$  is 102.22, which is roughly twice the ratio of edges to nodes. Finally, the clustering coefficient in  $G^W$  is 0.71. Figure 3.7 shows a plot of the cumulative degree distribution along with the best fitting exponential distribution. The exponential fit appears to be good in the lower half of the support, but the empirical distribution is skewed toward the upper tail.

Curiously, the ratio of edges to vertices and average degree for  $G^W$  are strikingly similar to those reported in table 3.2 for the network of film actors reported by Newman (2003). They share very similar ratio of edges to nodes, average degree, and clustering coefficient. Both networks are formed by projecting a network linking individuals and their employers onto the set of individuals. However, a crucial topological difference is that the degree distribution of the film star network follows a power law distribution. In terms of the processes generating the network data, the major distribution is that film actors are guaranteed to have worked on the same film at the same time, while connection in the realized mobility network does not imply that two people worked together, or even that they had the same type of job at the same time; only that they share a common element in their job histories. As was discussed previously, the exponential distribution of degree in  $G^W$  may be interpreted as evidence of substantial randomness in the connections between workers on the basis of shared jobs. Conversely, the power law distribution of degree in the film star network suggests non-random connections between movie stars on the basis of shared films. This interpretation seems intuitively plausible: famous movie stars will tend to appear in more films, and to appear in more films with each other, while more obscure actors are in fewer films, but tend to be in these together.

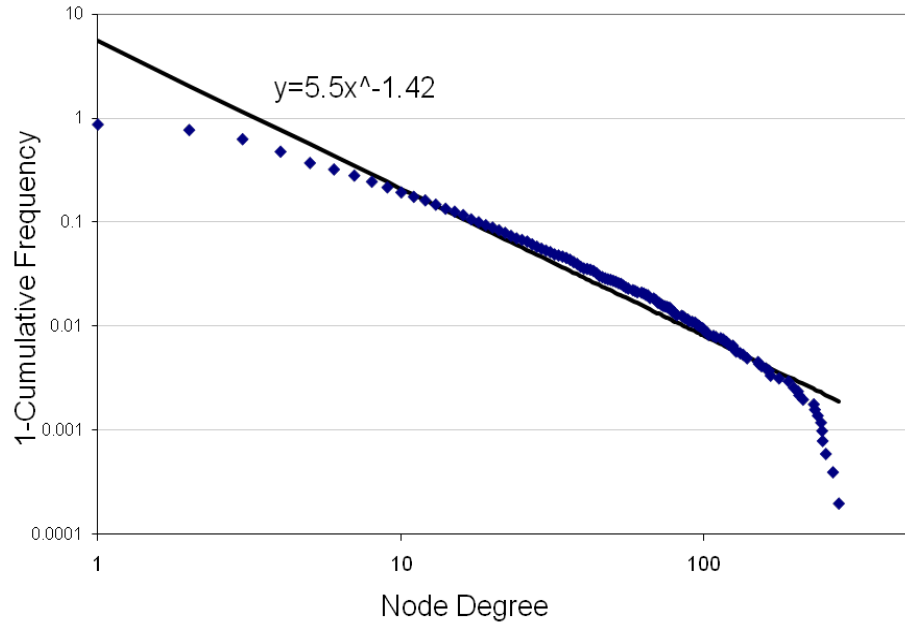


Figure 3.8: Degree distribution of the projection of the realized mobility network in the PSID sample onto pseudo-employers

### The Projection onto $M : G^M$

$G^M$  is quite a bit more sparse than  $G^W$ , most likely because of the restriction on the number of jobs that any given worker can link together. In  $G^M$ , 26,614 edges connect the 5,255 pseudo-employers. Naturally, there is also only one large component in  $G^M$  containing 4,999 of the nodes. The average degree is 9.5, again roughly twice the ratio of edges to nodes. Clustering in  $G^M$  is 0.57. Figure 3.8 shows a plot of the degree distribution, along with the best fitting power law distribution. The power law distribution looks like a reasonable fit in the middle of the support, but underpredicts in the tails; a fairly standard result. The only network in the literature I reviewed that had a similar topology is the network of coauthorship relations between academic physicists (Newman 2003). This similarity is somewhat harder to motivate than the similarities be-

tween  $G^W$  and the film star network, and since I lack information on the shape of the degree distribution in the physics coauthorship network, it is hard to draw many implications from the similarity.

The fact that  $G^M$  follows a power law is quite interesting when contrasted with the shape of the degree distribution in  $G^W$ . The connections between pseudoemployers based on shared employees has non-random structural features, while the connections between workers based on shared jobs are essentially random. This parallels a finding by Newman et al. (2001) who analyzed the bipartite graph between company directors and the boards on which they serve; the so-called board-interlock network. That study finds that the degree distribution in the one-mode projection onto the set of boards shows randomness, but that the degree distribution in the projection onto the directors is rather non-random.

These results also agree with the analysis in Guillaume and Latapy (2004) who argue that complex bipartite networks tend to have a power law on one set of nodes. They also show how the form of the degree distribution will be preserved through the projection onto one or the other set of nodes. The realized mobility network is consistent with a version of their random growth model in which the number of jobs held by a worker over his life is a Poisson random variable, but the employers with which these jobs are held are chosen by preferential attachment. These abstract features are consistent with conventional economic models in which job search behavior of workers is conditionally random, but workers are more likely through this random mobility to move into pseudoemployers that are large. Here, workers are more likely to move into industries and occupations that have many other employees.

### 3.6 Finding Mobility Clusters using Community Detection Across Workers

The intuition behind using the realized mobility network to think about market clusters can be neatly motivated using an example based on labor-market segmentation. Imagine that the vertex set of the realized mobility network,  $V(G)M$ , is partitioned into two non-intersecting segments,  $(W_1, M_1)$  and  $(W_2, M_2)$ . Think, for instance, of  $M_1$  as the set of high-wage employers and  $M_2$  as the set of low-wage employers. We might imagine that there are two sets of workers,  $W_1$  and  $W_2$ , that are stuck working for low-wage employers in the following sense: worker  $i$  has probability  $p_{in}$  of getting a job with an employer in his own set of the partition, and probability  $p_{out}$  of connecting to a job in the opposite set. If this is so, then depending on the level of mobility, there will be more edges connecting workers to multiple employers in the same segment than there are edges connecting workers to employers in the opposite segment.

In the extreme case where the probability of transition between the two segments is zero, the realized mobility network will always have two distinct components corresponding to the two segments. If this were the case, then we could check for the existence of different mobility segments by simply looking for unconnected components in the graph. Even without turning to the data, the results from complex network analysis tell us not to expect this method to work. Complex networks are almost universally characterized by having one large connected component (Newman 2003). While the realized mobility network is not a social network in the pure sense, if anything it will have greater connectivity than a standard social network. Indeed, the realized mobility network in



Unemployment Insurance data from the LEHD analyzed in Abowd et al. (2002) has one large component that includes most of the nodes in the network. We therefore know that finding market segments will be more complicated.

When the probability  $p_{out} > 0$ , the underlying partition is not revealed by the component structure of the graph. Instead, we will have a graph in which there is one large connected component containing nodes from both  $(W_1, M_1)$  and  $(W_2, M_2)$ . However, it should be the case that within each set of the partition, the connections are more dense. The empirical problem, which can be addressed using tools of complex network analysis, is how to identify the partition by identifying those pockets of relatively higher density. This is known in the literature on complex networks as a problem of finding community structure.

Clauset et al. (2004) develops a new method for finding community structure based on the concept of graph modularity. This is implemented through their FastCommunity algorithm, which produces results equivalent to those found using the earlier algorithm, but has linear run times for most graphs, so that community structure is computable in graphs with millions of vertices.

Modularity is an objective function whose domain is the set of possible partitions of  $W$  and whose value depends on the structure of  $G^W$ . To formalize this a bit, let  $\Pi(W)$  be the set of all partitions of  $W$ . So a representative element  $\pi = (\pi_1, \dots, \pi_K) \in \Pi(W)$  is a division of  $W$  into  $K$  mutually exclusive and exhaustive subsets. In the current setting,  $\pi$  is a proposed division of  $W$  into communities. The modularity of  $G^W$  is meant to measure the extent to which the proposed division fits the structure of connections in the graph better than it would fit a graph in which the connections were purely random. If  $G^W$  were

connected completely at random with identical probability of connection between each pair of vertices, then a proposed partition  $\pi$  will combine densely connected sets of vertices only by chance.

To define modularity formally, we need some additional terminology.

Let  $A^W$  be the adjacency matrix of  $G^W$ . Given  $v, w \in W$ ,  $A_{vw}^W = 1$  if  $vw$  is an edge in  $G^W$  and is zero otherwise.

Let  $m$  be the number of edges in  $G^W$

Let  $\delta_\pi : W \times W \rightarrow \{0, 1\}$  be a function such that  $\delta_\pi(v, w) = 1$  if there exists  $\pi_k \in \pi$  such that  $v \in \pi_k$  and  $w \in \pi_k$ . are in the same community in  $C(W)$  and is zero otherwise.

Let  $k_v$  be the degree of node  $v$ .

The goal now is to specify a function,  $Q : \Pi(W) \rightarrow \mathbb{R}$  whose value is maximized when we have found the community structure with best fit. Whatever specification we choose for this function will necessarily embed a set of assumptions about what defines the community structure of the graph. In a working paper Copic et al. (2009) establish a decision-theoretic foundation for a maximum-likelihood type of objective function for finding community structure that is somewhat similar, but not identical, to modularity. Both that function and modularity suffer from the high dimensionality of the domain space, which is exponential in the number of edges. For this reason, true maximization is generally impossible and approximation methods are required.

Returning to the definition of modularity, note that given  $\pi$ , the fraction of

within-community edges in the graph is

$$\frac{1}{2m} \sum_{v,w \in W} A_{vw} \delta_{\pi}(v, w)$$

An alternative interpretation of this quantity is as the empirical probability that a given edge in the graph lies within a community. This is not a good candidate as an objective function, because it can be maximized using the trivial partition  $\pi = W$ . The idea of modularity is to consider the difference between this value and the empirical probability of a given edge lying within a community when all edges are formed with equal probability.

Taking the degree distribution of the graph as given, and assuming that all edges occur with equal probability, a basic combinatorial argument establishes that the probability that a randomly drawn edge does not connect two different communities is

$$\frac{1}{2m} \sum_{v,w \in W} \frac{k_v k_w}{2m} \delta_{\pi}(v, w)$$

The modularity of  $G^W$  given  $\pi$  is defined as

$$Q(\pi) = \frac{1}{2m} \sum_{v,w \in W} (A_{vw} - \frac{k_v k_w}{2m}) \delta_{\pi}(v, w)$$

The trivial partition,  $\pi = W$  gives  $Q(\pi) = 0$  while positive values for  $Q$  indicate the existence of a non-random community structure in the graph. In the tests reported in Clauset et al. (2004), a maximized value of modularity between 0.3 and 0.7 is indicative of significant community structure.

Table 3.3: Summary of Job Types in Coworker Segments

	Type of Jobs
Seg. 1	Blue Collar Industries; Managers, Foremen, Operators and Truck Drivers
Seg. 2	Health Care and Retail; Nurses, Managers and Clerical Workers
Seg. 3	School Teachers, Teaching Aides and Administrators
Seg. 4	Doctors, Lawyers, Bank Officers, College Professors
Seg. 5	Postal Workers, Firemen

Descriptions of highest frequency industries and occupations in each of the coworker segments.

### 3.6.1 Segment Structure

Applying the FastCommunity algorithm of Clauset et al. (2004) to the large component in  $G^W$  reveals 21 separate segments at a maximal modularity of 0.48. Of these segments, the smallest 16 contain only 110 nodes. For the rest of the paper, I treat these 16 smallest segments as ‘segment 6’. The FastCommunity algorithm appears to break the network up in a non-random manner. That is, the six segments, which were detected on the basis of mobility patterns alone, have distinct demographic and economic characteristics as we would expect to find in different market segments.

I find that these segments have distinct industry and occupation profiles. Table 3.3 provides the titles of the dominant industries and occupations in each segment which groups of employers hire more frequently from each segment, or put another way, the pseudo-employers between which a worker is most likely to move. The largest two segments split largely along industry lines. Segment 1, containing approximately half of the nodes is concentrated in the blue-collar industries, while segment 2 is focused in the service sector, especially health

care. Both of these segments have heavy concentrations of managerial and non-managerial occupations. This phenomenon repeats in segment 3, which is concentrated in public education, but which includes teachers, teachers' aides, school administrators and principals. Thus, it appears to be the case that markets are separated by industry, but within industries there is relatively little occupational segmentation.

Segment 4 is the exception to this rule, and it is an interesting one. It has no distinct industrial concentration, but includes several occupations that one might think of as career occupations, such as doctors, lawyers, bank officials and college professors. On one hand, this has some intuitive appeal, as we would expect this group of workers to participate in a different market than the workers in segment 1-3.

### **3.6.2 Correlates of Worker Segment**

Table 3.4 shows the results from estimating a standard multinomial logit to determine the worker's segment on the basis of demographic and human capital characteristics. The model includes controls for age, number of jobs held, and labor market experience. Figures 3.9 and 3.10 show the predicted probability under the model of being in each segment conditional on being a white male and on levels of education. The correlation of race and gender on segment choice mostly conform to the underlying industrial and occupational composition. White males are much more likely to fall in segment 1, while non-white females are much more likely to fall in segment 2. Given the industrial composition of these two segments, this is an unsurprising result. Interestingly, race

Table 3.4: Logit model for Coworker Segment

	Seg. 1	Seg. 2	Seg. 3	Seg. 4	Seg. 5	Seg. 6
white	0.13* (0.015)	-0.10* (0.013)	-0.01 (0.008)	-0.00 (0.006)	-0.03* (0.005)	0.00 (0.003)
male	0.45* (0.011)	-0.28* (0.010)	-0.17* (0.008)	0.01 (0.005)	-0.02* (0.004)	0.01* (0.003)
High School	-0.07* (0.020)	-0.00 (0.017)	0.02* (0.013)	0.06* (0.016)	-0.00 (0.005)	0.00 (0.004)
Some College	-0.13* (0.214)	-0.11 (0.018)	0.03* (0.023)	0.11* (0.022)	-0.02* (0.005)	-0.01 (0.004)
College	-0.24* (0.022)	-0.00 (0.022)	0.15* (0.023)	0.12* (0.027)	-0.03* (0.004)	-0.00 (0.004)
Postgraduate	-0.36* (0.019)	-0.02 (0.023)	0.23* (0.030)	0.19* (0.035)	-0.03* (0.004)	0.00 (0.005)
Number of Obs.	7425					
Log-likelihood.	-8622.9931					

Marginal effects from a multinomial logit model predicting worker segment.

The model also controls for age, experience and number of job changes.

Standard errors in parentheses.

\*Statistically significant at the 2.5% level.

does not matter for entering segments 3 or 4, and even gender does not strongly predict membership in segment 4. Being female predicts being in segment 3, which is concentrated in the public schooling industry.

Increased education is a strong predictor of entering segments 3 and 4, which is what one would expect given the kinds of jobs workers in those segments hold. More highly educated workers are far less likely to be in segment 1, the blue-collar segment. Curiously, education is not a determinant of membership in segment 2, even though that segment has a concentration of health care professionals. This is especially interesting in light of the fact that workers in segment 2 and 3 are by far the worst paid, at least in terms of the median labor income.

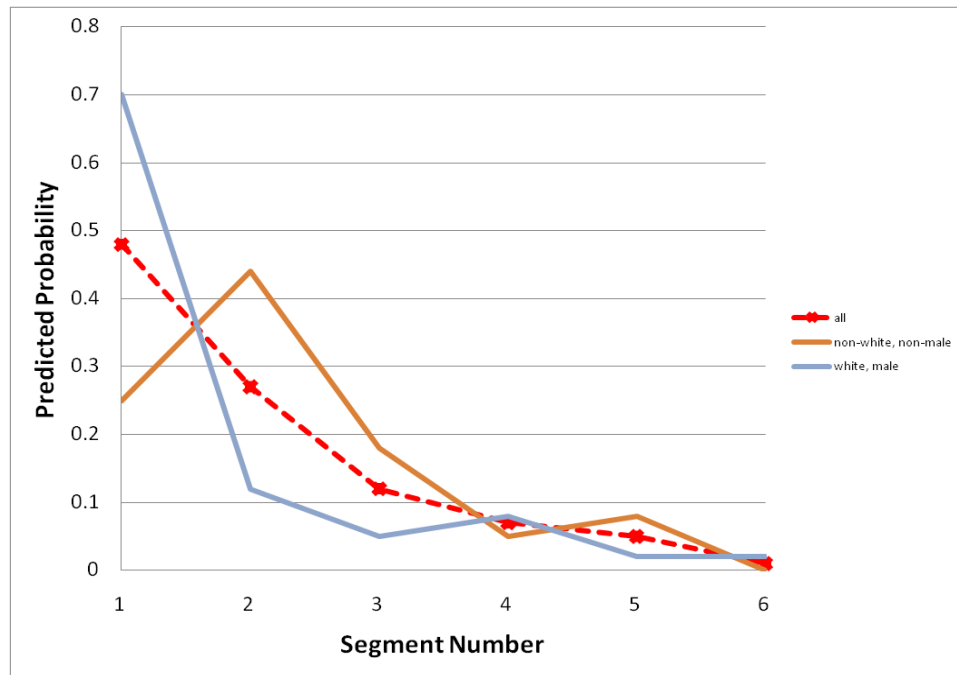


Figure 3.9: Predicted probability of appearing in each of the six segments found by maximizing modularity of the worker projection.

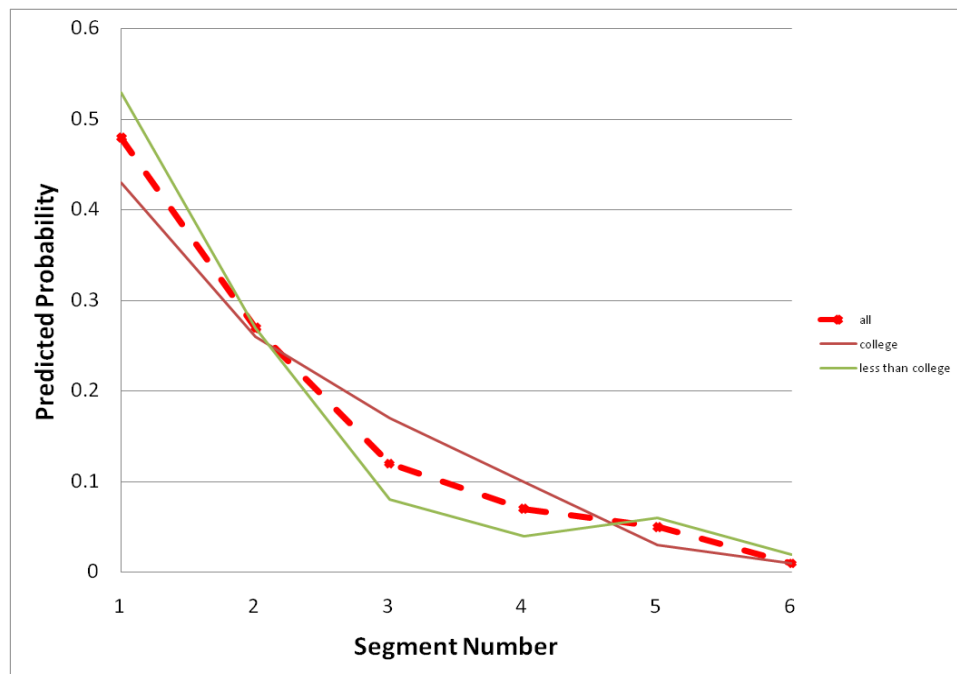


Figure 3.10: Predicted probability of appearing in each of the six segments found by maximizing modularity of the worker projection.

### **3.6.3 Wage Determination By Segment**

Interestingly, the segment of the labor market in which the worker participates does not seem to have a strong influence on wage determination. Table 3.5 shows the results of OLS regressions of the log wage on demographic and human capital variables, including controls for labor market experience and number of jobs held. The results are presented for a pooled regression as well as separately for workers within each segment. Contrary to the radical dual labor market hypothesis, there are returns to education and age within each segment, though these returns do differ quantitatively across segment. The influence of race and gender have the expected sign, except in segment 3 where there is no return to being white. Segment 2 also has a low, but significant return to being white. Without a formal model, it is difficult to interpret these patterns, but they do appear to suggest that at least some aspects of wage determination vary by segment.

## **3.7 Finding Mobility Clusters by Maximum Likelihood Partition of Employers**

### **3.7.1 A Basic Model Of Job Mobility**

This is a simple reduced-form model of job-mobility with dual labor markets. The baseline model assumes that mobility is determined only by the sector in which the worker is currently employed. In this baseline model, I have not incorporated the well-known fact that job mobility varies with age and tenure.



Table 3.5: Log Earnings Regression on Coworker Segments

	Seg.1	Seg.2	Seg.3	Seg.4	Seg.5	Pooled
white	0.211*	0.084*	-0.167*	0.171*	0.037	0.147*
	(0.0136)	(0.0180)	(0.0320)	(0.0288)	(0.0407)	(0.0095)
male	0.562*	0.539*	0.635*	0.622*	0.781*	0.703*
	(0.0132)	(0.0186)	(0.0354)	(0.0262)	(0.0421)	(0.0084)
number of jobs	0.025	-0.002	0.006	0.006	-0.011	0.013*
	(0.0022)	(0.0033)	(0.0055)	(0.0049)	(0.0078)	(0.0015)
age	0.012*	0.028*	0.027*	0.064*	0.056*	0.103*
	(0.0006)	(0.0025)	(0.0043)	(0.0052)	(0.0077)	(0.0036)
age <sup>2</sup>	-0.000*	-0.000*	-0.000*	-0.001*	-0.001*	-0.001*
	(0.0000)	(0.0000)	(0.0000)	(0.0001)	(0.0001)	(0.0000)
High School	0.335*	0.415*	0.505*	0.215*	0.322*	0.362*
	(0.0168)	(0.0269)	(0.0591)	(0.0671)	(0.0556)	(0.0133)
Some College	0.498*	0.640*	0.606*	0.425*	0.533*	0.539*
	(0.0181)	(0.0282)	(0.0596)	(0.0672)	(0.0621)	(0.0140)
College	0.823*	1.016*	1.181*	0.573*	0.596*	0.816*
	(0.0209)	(0.0324)	(0.0588)	(0.0686)	(0.0735)	(0.0155)
Postgraduate	1.018*	1.134*	1.523*	0.961*	0.390*	1.025*
	(0.0261)	(0.0356)	(0.0589)	(0.0684)	(0.1117)	(0.0169)
Intercept	8.367*	8.244*	7.724*	7.883*	7.832*	6.528*
	(0.0327)	(0.0550)	(0.1021)	(0.1260)	(0.1678)	(0.0769)
Number of Obs.	23,840	12,704	6,135	3,569	2,079	48,173
$\bar{R}^2$	0.1889	0.2048	0.2510	0.3044	0.2257	0.2366
Pooled Regression†	-0.094*	-0.190*	-0.534*	0.107*	-0.086	
	(0.0354)	(0.0359)	(0.0369)	(0.0379)	(0.0402)	
Number of Obs.	49,021					
$\bar{R}^2$	0.2409					

Description of the effects of segment affiliation on log annual labor market earnings. The Data are worker-year observations from the PSID sample. The segment affiliation of each worker-year observation is assigned based on the reported industry-occupation pair. The parenthesized values are standard errors. For purposes of comparison, the final column contains results from the same model estimated on the pooled sample.

\* statistically significant in the indicated direction at the 97.5% level of confidence.

† Estimates of the level effect associated with segment affiliation in a pooled regression including the same covariates where the effects of those covariates are restricted to be the same across segments.

On the one hand, this should not matter too much for the current application, since ultimately all that matters is the relative number of links within and across different partitions of the set of employers. On the other hand, if the relationship between tenure and mobility differs across sectors, this model could be misspecified in a way that affects our inference of the partition structure.

We have data on job histories of the form,  $D = \{(m, w, t)\}$ , where

- $m \in M$  is an employer (or pseudo-employer )
- $w \in W$  is a worker
- $t \in T$  denotes the time of the observation

In general, there is also information on wages, worker characteristics and employer characteristics. Each of these, especially wages, should be incorporated into a fully specified mobility model. Again, since the goal here is to identify segments solely on the basis of mobility patterns, I am ignoring that other information for the moment.

## Jobs

*Definition 5* A triple  $(m, w, t)$  is an element of the work history of  $w \in W$ . It says that  $w$  was employed by  $m$  at time  $t$ . A **job** is a collection of time-contiguous work-history elements for worker  $w$  with the property that all elements include the same employer.

A worker,  $w$ , will have  $k_w$  jobs during the sample. The worker's job history will look something like this:

- $j(w, 1) = \{(m_1, w, t_1), (m_1, w, t_1 + 1), \dots, (m_1, w, t_{j_1})\} = (m_1, t_1, t_{j_1})$
- $j(w, 2) = \{(m_2, w, t_{j_1} + 1), \dots, (m_2, w, t_{j_2})\} = (m_2, t_{j_1} + 1, t_{j_2})$
- ...
- $j(w, k_w) = \{(m_{k_w}, w, t_{j_{k_w}-1} + 1), \dots, (m_{k_w}, w, t_T)\} = (m_{k_w}, t_{j_{k_w}} + 1, t_T)$

Let  $J^w$  denote the complete job history of  $w$ , and let  $J^{w,t}$  be the partial history of jobs that is completed as of time  $t$ .

With the additional restriction that  $m_i \neq m_{i+1}$  for all  $i$ . Note that the time periods do not overlap. This is equivalent to assuming that workers only have one job at a time, or that only their main jobs are of any consequence. Also, the time periods for each job are directly adjacent, so there is never an instance where a worker is not in a job. If we enlarge the concept of a job to include states of unemployment and transitions out of the sample, then this is not a problem. Finally, note that the time period at which the first job begins and the last job ends are censored. For the model I develop here, this does not really matter, but in general it will.

## Job and Employer Types

To introduce the idea of labor market segmentation, we assume that each job has a latent type,  $\omega_j \in \{0, 1\}$  where  $\omega_j = 1$  if  $j$  is a primary sector job and  $\omega_j = 0$  if  $j$  is a secondary sector job. In general, a single employer  $m$  can offer jobs of both types, just as a single worker can hold jobs of both types. For the purposes of this paper, assume that employers only ever offer jobs of one type or the other. In this case it makes sense to assert that each employer  $m$  has a constant type,

$\theta_m \in \{0, 1\}$  defined just like  $\omega_j$ . This is the latent variable generating variation in mobility rates. The goal is to estimate this variable.

## Job Mobility

Define  $\mu(w, t) = m$  as the function that maps a worker and time period onto the worker's employer during that period. We are interested in the way workers move between employers. As a very general statement of the problem, we are interested in the probability that a worker  $w$  is employed by employer  $m$  conditional on the worker's entire previous job history as well as the job histories of all other workers. That is, the probability that  $w$  is employed by  $m$  at  $t$  might depend on everything that has happened in the labor market up to that point.

$$p_{n,t}^w = \Pr(\mu(w, t) = n | J^{w,t-1}, J^{-w,t-1})$$

The basic mobility model pursued here is substantially simpler. In particular, I make the following assumptions:

1. (Markov)  $p_{n,t}^w = \Pr(m(w, t) = n | m(w, t-1) = m)$ . Where you work in the next period depends only on where you work currently.
2. (Conditional Homogeneity)  $p_{n,t}^w = p_{n,t}^{w'}$  for all  $w, w' \in W$ . Job mobility does not depend on who you are; only where you just worked.
3. (Stationarity)  $p_{n,t}^w = p_{n,t'}^w$  for all  $t, t' \in T$ .

Therefore, it will make more sense for us to adopt the notation

$$p_{mn} = \Pr(\mu(w, t) = n | \mu(w, t-1) = m)$$

## Labor Market Segmentation

The model of labor market segmentation allows for even more simplification of the mobility model.

At the beginning of each time period,  $t$ , each worker experiences a stochastic job history event,  $s_{w,t} \in \{s_1, s_2, s_3\}$  where

- if  $s_1$  occurs, the worker stays with his current employer  $m$
- if  $s_2$  occurs, the worker switches to a different primary sector employer
- if  $s_3$  occurs, the worker switches to a different secondary sector employer

The hypothesis that there is limited mobility between two segments is formalized as follows. Let

$$q_{lm} = \Pr(s_{w,t} = s_m | \theta_{\mu(w,t)} = l)$$

That is, the probability that the worker draws a particular job history event depends on the state of his current employer. If there are no mobility effects of labor market segmentation, then we should have  $q_{l2} = q_{l3}$  for  $l = 0, 1$ . The mobility restrictions in the segmentation model are formalized through the assumptions:

$$q_{03} > q_{02}$$

$$q_{12} > q_{13}$$

These say that a worker is more likely to move to an employer of the same

type than one of the other type. A weaker assumption is:

$$q_{01} + q_{03} > q_{02}$$

$$q_{11} + q_{12} > q_{13}$$

Which says that the probability that you add another unit of tenure within your sector is greater than the probability of changing sector. The first assumption can be thought of as conditioning on there having been a move. In general, we expect  $q_{01} > q_{02} + q_{03}$  and analogously for the primary sector.

We complete the mobility model under segmentation by assuming that once we know  $s_{w,t}$ , the worker moves to a randomly drawn employer. The probability of the move is given by  $\eta_{m,t}^\theta$  where  $\eta_{m,t}^\theta = 0$  if  $\theta_m \neq \theta$ . (There are two probability measures over the space of employers; one assigns positive probability only to workers in the primary segment and the other only to workers in the secondary segment). So, we finish the mobility model by noting that

$$p_{mn} = \begin{cases} q_{\theta_m,1} & \text{if } m = n \\ q_{\theta_m,2} \left( \frac{\eta_{n,t}^1}{1-\eta_{m,t}^1} \right) & \text{when } \theta_n = 1 \\ q_{\theta_m,3} \left( \frac{\eta_{n,t}^0}{1-\eta_{m,t}^0} \right) & \text{when } \theta_n = 0 \end{cases}$$

Finally, assuming that once a worker moves, the employer is chosen from the set of all firms of the same type with equal probability, (that is,  $\eta_{m,t}^\theta = \eta_{m',t'}^\theta$  for all  $m, m', t, t'$ ), we can write this simply as

$$p_{mn} = \begin{cases} \pi_{s,\tau\tau} & \text{if } m = n \text{ and } \theta_m = \tau \\ \pi_{s,\tau\tau'} & \text{when } m \neq n \text{ and } \theta_m = \tau \text{ and } \theta_n = \tau' \end{cases}$$

So, the final model is characterized by six probabilities associated with different kinds of job history events along with the latent variable  $\theta$  describing

employer types.

### 3.7.2 The Structure of Employer-Employer Links Predicted in the Model

Consider a graph defined from the realized mobility network being projected onto the set of employers,  $M$ . We denote this graph by  $G^M$  and make the following definitions:

- The vertex set of  $G^M$ ,  $V(G^M) = M$
- The edge set of  $G^M$ ,  $E(G^M)$  is derived as follows.  $(m_1, m_2) \in E(G^M)$  whenever there is  $w \in W$  and  $t \in T$  such that  $(m_1, w, t) \in D$  and  $(m_2, w, t + 1) \in D$ . That is, we observe some worker moving from  $m_1$  to  $m_2$ .

Note that  $G^M$  is a directed graph allowing both multi-edges and self-edges. It is therefore different from the graph labelled  $G^M$  in section 3.2. Intuitively, this graph has an edge for every worker that made a transition between those two employers. Furthermore, a node has a self-edge for every period and every worker that stayed with that employer. Therefore, the edges in the network contain information about mobility between employers, and about tenure with employers.

Following Copic et al. (2009), let  $g_{mn}$  be the number of observed links from  $m$  to  $n$ , and let  $s_{mn}$  be the number of potential links between  $m$  and  $n$ . Intuitively,  $s_{mn}$  should be equal to the sum of number of workers who are ever at  $s_{mn}$  times

the number of time periods they are there, since this is the number of chances for a worker to start at  $m$  and move to  $n$ .

With this notation in mind, and given the independence assumptions in the model above, observe that the chance of seeing exactly  $g_{mn}$  moves from  $m$  to  $n$  is

$$p_{mn}^{g_{mn}} (1 - p_{mn})^{s_{mn} - g_{mn}}$$

Where  $p_{mn}$  is substituted from the job mobility model above.

We can regard  $G^M$  as having been generated by the job mobility model outlined above. In this case, we can write the likelihood function:

$$\begin{aligned} \Pr(\theta, \pi | G^M) &\propto \prod_{m,n} p_{mn}^{g_{mn}} (1 - p_{mn})^{s_{mn} - g_{mn}} \\ &= \prod_{m,n} \left[ \pi_{s,00}^{g_{mn}} (1 - \pi_{s,00})^{s_{mn} - g_{mn}} \right]^{\mathbf{1}(m=n, \theta_m=0)} \left[ \pi_{s,11}^{g_{mn}} (1 - \pi_{s,11})^{s_{mn} - g_{mn}} \right]^{\mathbf{1}(m=n, \theta_m=1)} \\ &\quad \left[ \pi_{m,00}^{g_{mn}} (1 - \pi_{m,00})^{s_{mn} - g_{mn}} \right]^{\mathbf{1}(m \neq n, \theta_m=0)} \left[ \pi_{m,00}^{g_{mn}} (1 - \pi_{m,00})^{s_{mn} - g_{mn}} \right]^{\mathbf{1}(m \neq n, \theta_m=1)} \\ &\quad \left[ \pi_{m,01}^{g_{mn}} (1 - \pi_{m,01})^{s_{mn} - g_{mn}} \right]^{\mathbf{1}(\theta_m=0, \theta_n=1)} \left[ \pi_{m,10}^{g_{mn}} (1 - \pi_{m,10})^{s_{mn} - g_{mn}} \right]^{\mathbf{1}(\theta_m=1, \theta_n=0)} \end{aligned}$$

Taking logs yields

$$\begin{aligned} \ln \Pr(\theta, \pi | G^M) &\propto T_{s,00}(g) \gamma_{s,00} + (T_{s,00}(s) - T_{s,00}(g)) \sigma_{s,00} \\ &\quad + T_{s,11}(g) \gamma_{s,11} + (T_{s,11}(s) - T_{s,11}(g)) \sigma_{s,11} \\ &\quad + T_{m,00}(g) \gamma_{m,00} + (T_{m,00}(s) - T_{m,00}(g)) \sigma_{m,00} \\ &\quad + T_{m,11}(g) \gamma_{m,11} + (T_{m,11}(s) - T_{m,11}(g)) \sigma_{m,11} \\ &\quad + T_{m,01}(g) \gamma_{m,01} + (T_{m,01}(s) - T_{m,01}(g)) \sigma_{m,01} \\ &\quad + T_{m,10}(g) \gamma_{m,10} + (T_{m,10}(s) - T_{m,10}(g)) \sigma_{m,10} \end{aligned}$$



where  $T_{s,00}(g)$  is the total number of self-links for employers of type 0 in the graph and  $T_{s,00}(s)$  is the total capacity for such links. Similarly,  $T_{m,01}(g)$  is the total number of edges in the graph from type 0 nodes to type 1 nodes and  $T_{m,01}(s)$  is the total capacity of such nodes.  $\gamma_{s,00} = \ln(\pi_{s,00})$ .  $\sigma_{s,00} = \ln(1 - \pi_{s,00})$ . The rest of the definitions are equivalent. This notation follows Copic et al. (2009).

Determining the segment structure of the labor market is equivalent to choosing the parameter  $(\hat{\theta}, \hat{\pi})$  that maximizes the above expression. There are two complications here: first, we have to deal with the fact that the choice of the optimal  $\hat{\theta}$  depends on  $\hat{\pi}$  and vice-versa. Second, this is a high-dimension discrete choice problem. We cannot search through the entire space  $\Theta = \{0, 1\}^{|M|}$ , so some approximation method is required.

Having obtained an estimate of  $\theta$ , the next step is to formally test whether it is substantially different from either  $\theta = (1, 1, \dots, 1)$  or  $\theta = (0, 0, \dots, 0)$ . That is, I want to formally test whether there are two distinct groups of employers between which mobility is relatively unlikely. Since this model is nested, it can be estimated using a likelihood ratio test.

### 3.7.3 Maximizing the Log-Likelihood

Now using the notation of graph theory becomes useful. Let  $A$  be the adjacency matrix for  $G^M$ . So  $A_{mn} = \#$  of links between  $m$  and  $n$ . Similarly, define  $S$  to be the  $|M| \times |M|$  matrix whose entries are the capacities of each edge. Recall that  $\theta$  is the  $|M| \times 1$  vector describing the sectoral affiliation of each employer,  $m$ . The

following relations are instructive:

$$\begin{aligned}
T_{s,00}(g) &= \sum_m (1 - \theta_m) A_{mm} \\
T_{s,11}(g) &= \sum_m \theta_m A_{mm} \\
T_{m,00}(g) &= \sum_{m \neq n} (1 - \theta_m)(1 - \theta_n) A_{mn} \\
T_{m,11}(g) &= \sum_{m \neq n} \theta_m \theta_n A_{mn} \\
T_{m,01}(g) &= \sum_{m \neq n} (1 - \theta_m) \theta_n A_{mn} \\
T_{m,10}(g) &= \sum_{m \neq n} \theta_m (1 - \theta_n) A_{mn}
\end{aligned}$$

we can rewrite the log-likelihood,  $\ln \Pr(\theta|G^M, \pi) =$

$$\begin{aligned}
& \sum_m \left\{ A_{mm} [\theta_m ((\gamma_{s,11} - \sigma_{s,11}) - (\gamma_{s,00} - \sigma_{s,00})) + (\gamma_{s,00} - \sigma_{s,00})] \right. \\
& \quad \left. + S_{mm} [\theta_m (\sigma_{s,11} - \sigma_{s,00}) + \sigma_{s,00}] \right\} \\
& + \sum_{m \neq n} A_{mn} \left\{ \begin{aligned} & \theta_m \theta_n \left[ \begin{aligned} & (\gamma_{m,11} - \sigma_{m,11}) - (\gamma_{m,00} - \sigma_{m,00}) \\ & - (\gamma_{m,01} - \sigma_{m,01}) - (\gamma_{m,10} - \sigma_{m,10}) \end{aligned} \right] \\ & + \theta_m [(\gamma_{m,10} - \sigma_{m,10}) - (\gamma_{m,00} - \sigma_{m,00})] \\ & + \theta_n [(\gamma_{m,01} - \sigma_{m,01}) - (\gamma_{m,00} - \sigma_{m,00})] \\ & + [(\gamma_{m,00} - \sigma_{m,00}) + (\gamma_{m,01} - \sigma_{m,01}) + (\gamma_{m,10} - \sigma_{m,10})] \end{aligned} \right\} \\
& + S_{mn} \left\{ \begin{aligned} & \theta_m \theta_n [\sigma_{m,11} - \sigma_{m,00} - \sigma_{m,10} - \sigma_{m,01}] + \theta_m [\sigma_{m,10} - \sigma_{m,01} - \sigma_{m,00}] \\ & + \theta_n [\sigma_{m,01} - \sigma_{m,10} - \sigma_{m,00}] + (\sigma_{m,00} + \sigma_{m,10} + \sigma_{m,01}) \end{aligned} \right\}
\end{aligned}$$

Evaluating the above expression is straightforward given a proposed partition, but the search space is both discrete and of large dimension, making this problem an ideal candidate for simulated annealing.

Table 3.6: Probability of Transition Between Pseudoemployer Segments

$q$	Estimate	Description
$q_{0t}$	0.2987	Stay in type 0
$q_{00}$	0.4374	Move from 0 to 0
$q_{01}$	0.2639	Move from 0 to 1
$q_{1t}$	0.5734	Stay in type 1
$q_{10}$	0.1435	Move from 1 to 0
$q_{11}$	0.2831	Move from 1 to 1

### 3.7.4 The Likelihood-Maximizing Partition

The best solution,  $\hat{\theta}_{MLE}$  was found by simulated annealing.  $\hat{\theta}_{MLE}$  puts 489 pseudoemployers into segment 1 – roughly 10% of the nodes.

The segmentation model implies that the within-segment transitions should be more likely than between-segment transitions. From  $\hat{\theta}_{MLE}$ , we compute these probabilities  $\hat{q}$ , which are reported in Table 3.6. As the segmentation model predicts, the probability of moving to an employer of the same type is higher than the probability of moving to an employer of the opposite type. Note also that the probability of staying is nearly twice as high in the small segment, segment 1. But, conditional on moving, the probability of leaving your segment is roughly 1/3 the probability of staying in it. It seems odd, however, that the probability of moving from a type 0 to another type 0 job is higher than the probability of staying in the same job in that segment. This finding could be explained by data errors. The erroneous coding of industry or occupation will show up in these data as spurious moves. So, it may be that some of the transitions we observe

should actually be recoded as accumulation of tenure in the same job. In either case, it remains true that leaving the segment is far less likely than staying within it.

We can use the Likelihood-Ratio test to show that the data reject the null hypothesis that there is no segmentation in the labor market in favor of the alternative that there are two segments. The null hypothesis is formalized as the assumption that  $\theta_0 = (0, 0, \dots, 0)$  is the likelihood maximizer. Under the null hypothesis, the test statistic

$$LR = -2[\log(L(A|\theta_0, q)) - \log(L(A|\hat{\theta}_{MLE}, q))] = 33,784$$

is distributed as  $\chi^2$ . The data reject the null hypothesis at any conventional level of significance.

### 3.7.5 Characteristics of the Segments

#### Industrial and Occupational Composition

Figures 3.11 and 3.12 show the industrial and occupational composition of the two pseudo-employer segments. These figures give the fraction of pseudo-employers in each one-digit industry or occupation within each segment. Because there are only two segments with this method, the separation is more coarse than in the previous section. Nevertheless, the results are largely consistent. Figure 3.11 gives the industrial composition, and shows that segment 1 is concentrated in service industries, while segment 0 is concentrated in manufacturing. Likewise, figure 3.12 indicates that segment 1 workers are more likely to be managers and service workers while segment 0 workers are more likely to

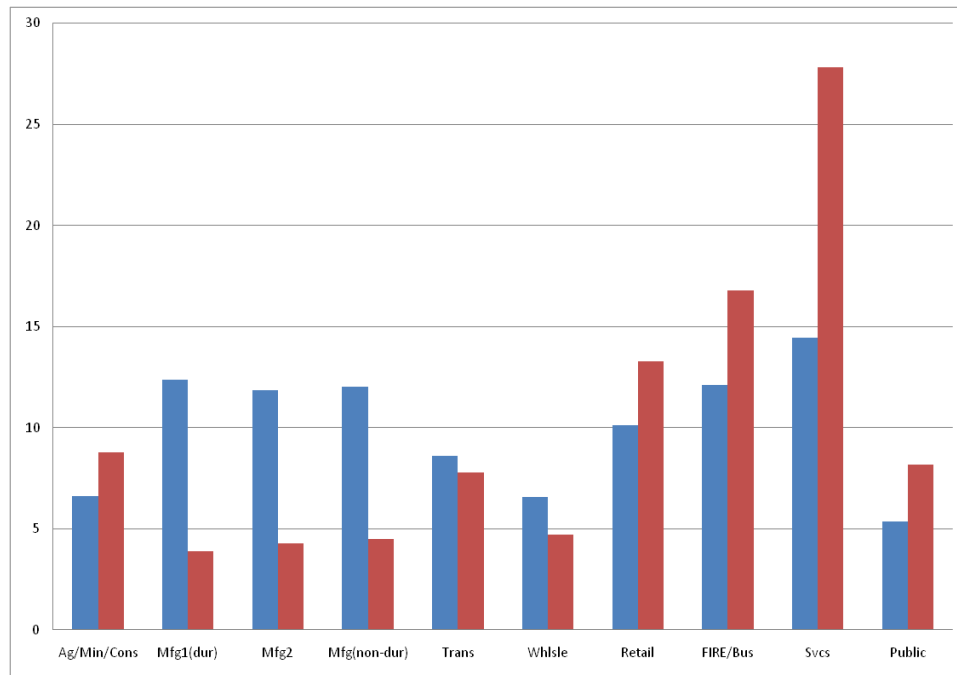


Figure 3.11: Industry Distribution by Segment. Blue is Segment 0, Red is Segment 1.

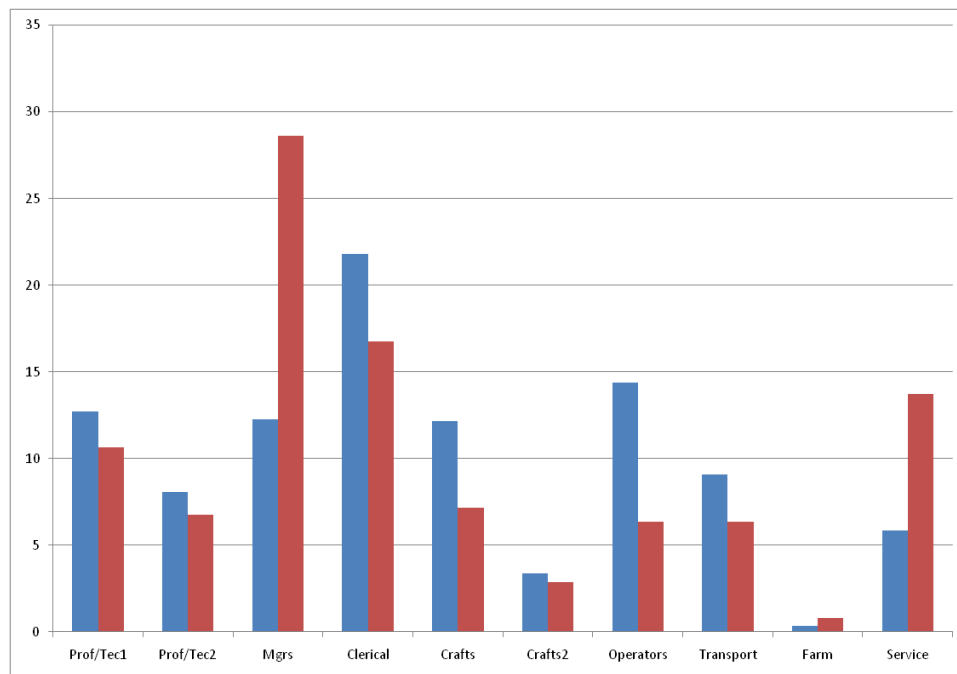


Figure 3.12: Occupation Distribution by Segment. Blue is Segment 0, Red is Segment 1.

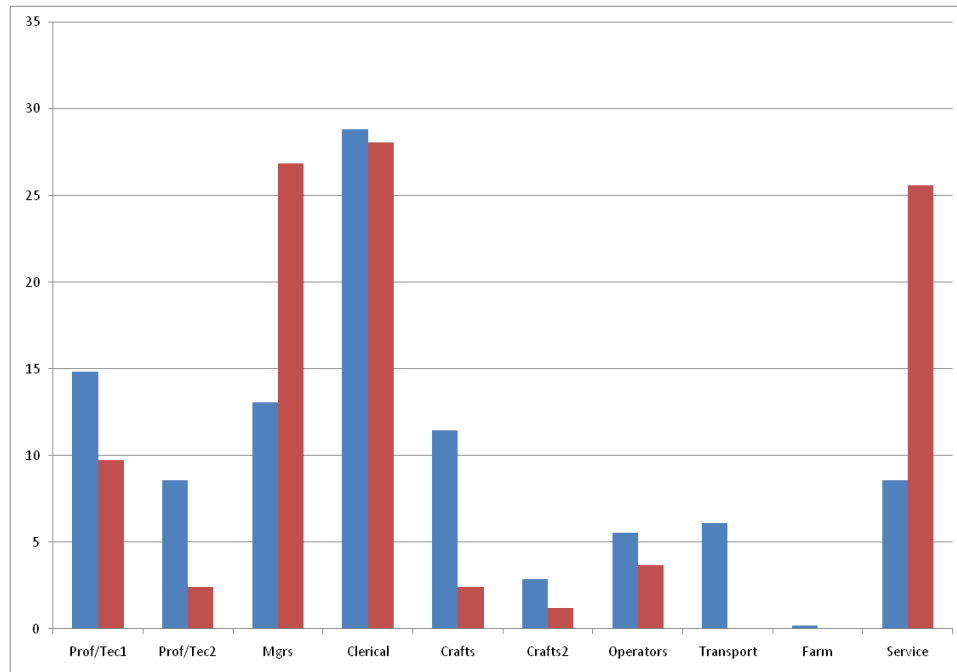


Figure 3.13: Occupation Distribution by Segment within FIRE. Blue is Segment 0, Red is Segment 1.

have blue-collar jobs.

Figures 3.13 through 3.15 show that there is still segmentation within industries, indicating that not all of the occupational segmentation detected with this method is associated with the occupational composition of particular industries. Within the FIRE industry, figure 3.13 shows that segment 1 jobs are more likely to be white collar, while segment 0 jobs are blue collar. Within manufacturing, the evidence is less clear, segment 1 jobs are more likely to be managerial than segment 0. There is also a division between service and clerical jobs in the Service sector. Altogether, these results suggest a division between white and blue collar jobs.

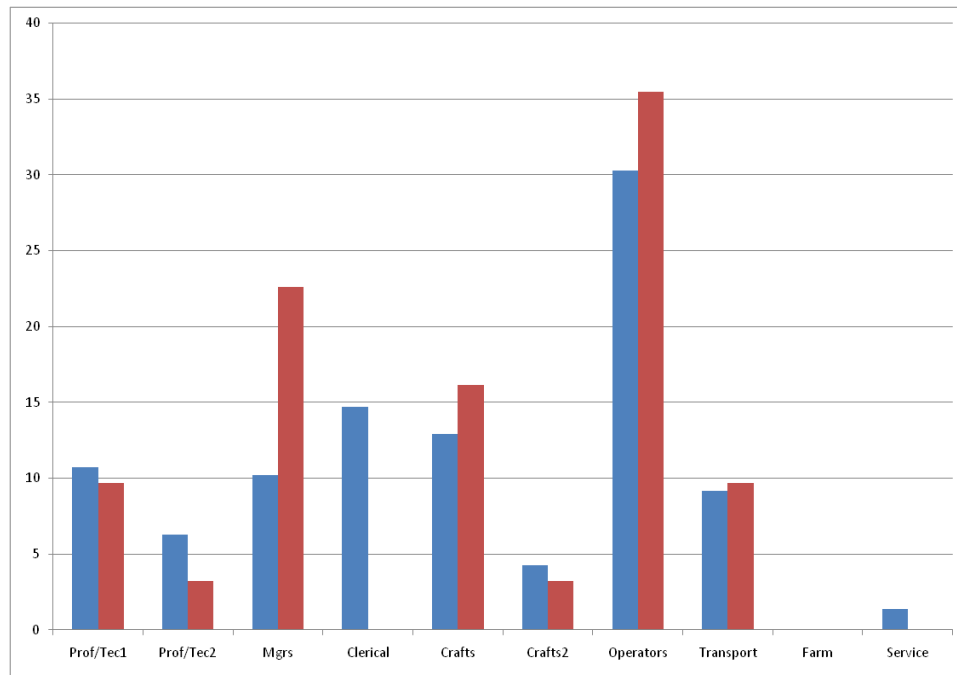


Figure 3.14: Occupation Distribution by Segment within Manufacturing. Blue is Segment 0, Red is Segment 1.

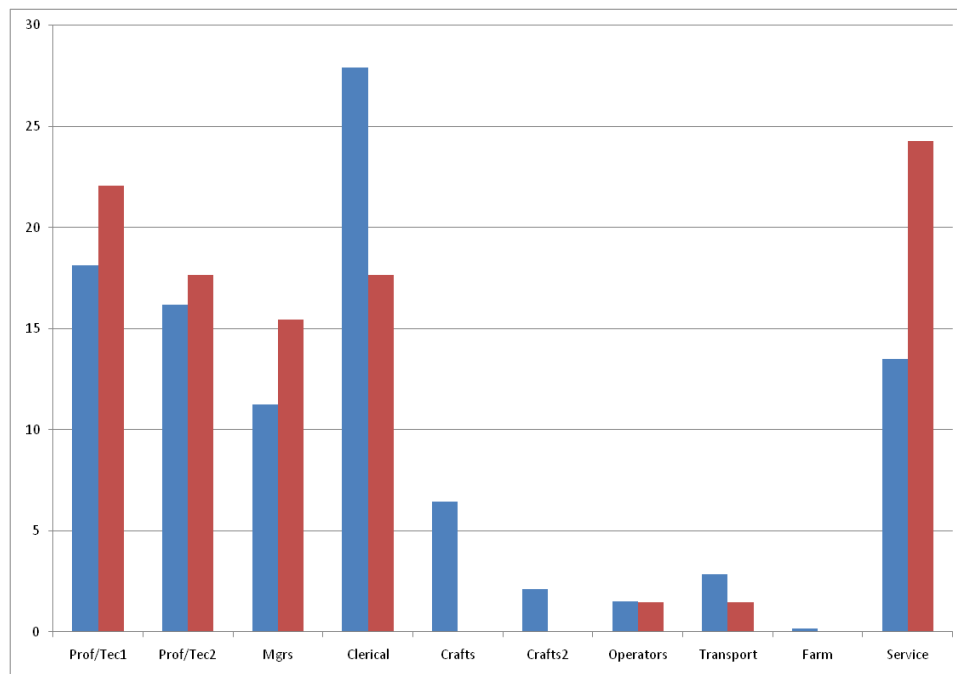


Figure 3.15: Occupation Distribution by Segment within the Service Sector. Blue is Segment 0, Red is Segment 1.

## Wage Determination

The relationship between labor market segments and wage determination is of longstanding interest. The dual labor market theory, for instance, predicts that workers in a 'secondary' segment will not earn returns to human capital, or at least not the same returns. Table 3.7 provides a descriptive analysis of the association between labor market segments and labor market earnings. Each column shows the result of a linear regression of the log of labor market earnings on individual characteristics, job characteristics, and segment affiliation. Model 1 is the baseline model, and does not control for any segment affiliation. The estimated coefficients are just as would be predicted in the literature: there are strong returns to human capital, a concave age-earnings profile, and positive wage premia associated with being white and male. Model 2 adds an indicator for segment 1. Recall that segment 1 jobs were more likely to be service jobs, but also more managerial. Without controlling for industry or occupation, being assigned a segment 1 job is associated with a -0.114 decrease in annual log earnings. However, we see in column 4 that controlling for the 1-digit industry and occupation of the job wipes out any significant segment effect while moderately attenuating the rest of the parameters.

Model 3 allows for arbitrary differences across segments in the coefficients on individual characteristics. So, if there were differences in returns to human capital across segments, we would expect to see it here. Model 3 shows that at least as a descriptive matter, there are differences across the segments. Jobs in segment 1 are associated with lower correlation between earnings and being white, and a higher correlation with being male, and having some education beyond high school. These findings persist when controlling for industry and



Table 3.7: Pseudoemployer Segment and Labor Market Earnings

	Model 1	Model 2	Model 3	Model 4	Model 5
white	0.144*	0.142*	0.205*	0.075*	0.147*
	(0.0189)	(0.0189)	(0.0251)	(0.0174)	(0.0235)
male	0.706*	0.693*	0.599*	0.608*	0.548*
	(0.0179)	(0.0179)	(0.0243)	(0.0192)	(0.0247)
age	0.104*	0.104*	0.102*	0.092*	0.089*
	(0.0682)	(0.0068)	(0.0094)	(0.0064)	(0.0088)
age <sup>2</sup>	−0.001*	−0.001*	−0.001*	−0.001*	−0.001*
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
High School	0.374*	0.376*	0.351*	0.262*	0.274*
	(0.0267)	(0.0265)	(0.0346)	(0.0252)	(0.0334)
Some College	0.559*	0.565*	0.472*	0.354*	0.330*
	(0.0285)	(0.0284)	(0.0386)	(0.0282)	(0.0378)
College	0.835*	0.847*	0.768*	0.517*	0.460*
	(0.0322)	(0.00320)	(0.0425)	(0.0349)	(0.0455)
Postgraduate	1.073*	1.090*	0.923*	0.746*	0.622*
	(0.0361)	(0.0360)	(0.0476)	(0.0398)	(0.0506)
Intercept	6.577*	6.641*	6.802*	7.397*	7.488*
	(0.1392)	(0.1394)	(0.1944)	(0.1388)	(0.1890)
Segment 1		−0.114*	−0.383	−0.004	−0.153
		(0.0139)	(0.2340)	(0.0144)	(0.2233)
Segment 1	white		−0.107*		−0.118*
			(0.0304)		(0.0284)
	male		0.148*		0.098*
			(0.0290)		(0.0275)
	age		0.005		0.004
			(0.0114)		(0.0110)
	age <sup>2</sup>		−0.0000		−0.0000
			(0.0001)		(0.0001)
	High School		0.044		−0.0176
			(0.0425)		(0.0403)
Segment 1	Some College		0.151*		0.041
			(0.0458)		(0.0434)
	College		0.132*		0.092
			(0.0523)		(0.0514)
Segment 1	Postgraduate		0.252*		0.180*
			(0.0575)		(0.0567)
Ind & Occ. Controls?	NO	NO	NO	YES	YES
R <sup>2</sup>	0.2392	0.2419	0.2446	0.2443	0.3123
Number of Obs.	45,221	45,221	45,221	45,221	45,221

Parenthesized values are standard errors clustered within individual. The bottom panel shows models that include an interaction of individual characteristics and the job being in segment 1.

\* statistically significant in the indicated direction at the 97.5% level of confidence.

occupation, as shown in model 5.

### **Predicting Segment Affiliation**

I use a logit model as a convenient method for summarizing the differences in the types of workers that have jobs in each segment. Using data on each new job, I estimate the segment of affiliation using individual characteristics. Model 1 in Table 3.8 shows that workers holding jobs in pseudo-employer segment 1 are more likely to be female and to be more highly educated than workers holding jobs in segment 0. These findings are consistent with segment 0 being concentrated in blue-collar jobs in manufacturing, while segment 1 jobs are more likely to be service jobs, and tilted toward white-collar occupations. A model that allowed for a larger number of segments might likely detect the cross-cutting industrial and occupational mobility influences that appear to be at work here.

## **3.8 Conclusion**

I have defined and examined an instance of the realized mobility network in a sample from the Panel Study of Income Dynamics (PSID). Analyzing the topology of the realized mobility network reveals several features of job mobility that call for closer study. First, the degree distribution and other static features of the network topology suggest that there is non-randomness in network formation, even in the absence of information on individual or pseudo-employer characteristics. These structural features could therefore be informative of otherwise unobservable characteristics of workers and employers in more conventional analyses of job matching.

Table 3.8: Factors Associated with Pseudoemployer Segment Assignment

	Model 1	Model 2
white	−0.020 (0.0265)	−0.019 (0.02817)
male	−0.3708* (0.02389)	−0.3335* (0.02701)
age	−0.001 (0.0013)	−0.000 (0.0014)
number of jobs	−0.156* (0.0047)	−0.150* (0.0051)
High School	0.121* (0.0347)	0.200* (0.0372)
Some College	0.2271* (0.0369)	0.3383* (0.0399)
College	0.400* (0.0428)	0.567* (0.0468)
Postgraduate	0.594* (0.0527)	0.843 (0.0579)
Log Earnings		−0.159* (0.0139)
Intercept	1.188 (0.0676)	2.524 (0.1355)
Number of Obs.	30,482	27,147
LR	1604.49	1708.71
Pseudo- $R^2$	0.0382	0.0456

Result of logit regression to predict the pseudoemployer segment of each new worker-pseudoemployer match.

\* statistically significant in the indicated direction at the 97.5% level of confidence.

Second, using two different methods for partitioning the realized mobility network, I found significant evidence of clustering among workers and pseudo-employers on the basis of observed job mobility. These mobility segments are distinct in terms of incomes, demographic characteristics, and job types. Both coworker segments discovered by maximizing modularity(method 1) and the pseudo-employer segments discovered by maximum likelihood (method 2)

show segment structure associated with occupational and industrial mobility, but with complex patterns that may not be revealed in a conventional analysis of inter-occupational or inter-industry transition rates. For example, the coworker segments show that workers in manufacturing and service industries tend to stay within those industries, suggesting that industry may be more influential than occupation in determining mobility, but the opposite appears to be the case for workers with substantial occupation-specific skills.

Both methods of segmentation generated non-random splits of the data in terms of individual characteristics and earnings. They also show that segmentation has a strong correlation with the estimated correlation between individual characteristics and earnings, though not in the extreme form predicted by dual labor market theory. In the coworker segments, there is substantial variation in the estimated relationship between education, demographic characteristics, and earnings, as well as in the estimated . The same is true of the raw correlations associated with the pseudo-employer segments. In the latter case, it is possible to eliminate most of the variation between segments, but only after controlling for industry- and occupation- effects as well as unobservable individual heterogeneity in earnings. Thus, analysis of the realized mobility network was able to uncover differences in earnings associated with individual-level unobservables.

All of this supports the claim of the paper that the realized mobility network contains important and useful information for economic analysis. In this paper, I have gone to one extreme in analyzing the labor market solely in terms of realized mobility patterns. Even at this extreme, I was able to uncover distinctions in industry- versus occupation specific mobility and earnings patterns associated with unobserved human capital. The next steps in this research pro-

gram are to develop models of labor market matching as a process of network evolution that properly conditions on all available data, and to fit these models using population-level matched employer-employee data. This work will use data on workers, employers and jobs back into the analysis. Soderberg (2002) extends the conventional model of random graph formation to condition connection probabilities on node characteristics. It is possible to extend his formalism to statistical inference of complex networks with a finite number of node types. There is also a literature in sociology on longitudinal network analysis beginning with Holland and Leinhardt (1977) that develops agent-based models of networks formation in which links are formed on the basis of both node and edge characteristics, as well as on the basis of networks structure as a whole. Snijders (2005) and Koskinen and Snijders (2007) develop frequentist and Bayesian methods for fitting dynamic network models.

If such methods can be developed, there are a range of interesting and important questions in labor economics to which they could be applied. For instance, these methods may be useful in finding methods to correct for endogenous assignment of workers to firms when estimating the wage premium associated with having a particular employer. The same technique could be modified to address related questions dealing with the impact of teachers on student achievement, or the effect of social groups on individual productivity. These methods may also be useful for finding the boundaries of markets for particular types of labor, whether those boundaries be related to skill accumulation, or are geographic. This would help in applications that require detailed understanding of the extent of a local labor market.

## CHAPTER 4

### ENDOGENOUS MOBILITY (WITH JOHN M. ABOWD)

#### 4.1 Introduction

The objective of this paper is to explore the consequences for estimates of worker and firm effects in the decomposition of earnings in longitudinally-linked employer-employee data when there is endogenous mobility, and to propose methods of correcting those estimates. Abowd et al. (1999) pioneered the identification, computation, and inference for the fixed effect estimator of the decomposition of log earnings into components associated with unobserved worker and employer heterogeneity. A major factor in the interpretation of their statistical model is that it requires the assumption that the assignment of workers to firms is random conditional on all observable characteristics and the design of the stationary unobservables. This assumption is at odds with many models of job assignment, in particular, those in which workers sort into jobs according to their comparative advantage. Since structural interpretations of the measured heterogeneity have major consequences for our understanding of the labor market, it is important to weaken the assumption that job mobility and assignments are exogenous to earnings.

The central problem of this paper is one manifestation of a fundamental challenge of empirical social science: separating the influence of correlated unobservables and sorting from the direct effect of group membership. Our approach is analogous to estimating treatment effects in the presence of selection on unobservables. It is just that here the number of possible treatments, that is employers, numbers in the millions. We construct an instrument for the ac-

tual assignment of workers to firms that exploits the relational structure of our data. The key insight is that the work histories of one's coworkers and previous employers are informative of one's own employment history, while being plausibly unrelated to whatever idiosyncratic wage innovations drive assignment at the margin.

Correcting estimates firm effects on wages for endogeneity bias is useful in a number of applications. First, this paper contributes to the ongoing debate as to whether estimates of employer-specific wage premia constitute evidence in contradiction of the law of one wage. If the bias correction does not affect the significance of firm effects, it suggests that firms really do play an important direct role in wage determination, consistent with sociological evidence, but contrary to the neoclassical model of a competitive labor market. Furthermore, it helps to resolve some of the debates spawned by the early empirical results based on the assumption of exogenous mobility. These early results show little correlation between estimated employer wage premia and worker-specific earnings. In other words, high wage workers do not systematically appear in high-wage firms. This has often been cited as evidence against theories of assortative matching, and in favor of models of frictional search, which predict this lack of assortativity. The empirical results have spawned a theoretical literature attempting to construct frictional search models with assortative matching, in which estimated firm- and worker- wage components would misrepresent the true assignment structure in the economy (Abowd et al. 2009; Lentz Forthcoming).

Estimates of person- and firm-effects are also being increasingly used in downstream applications. Iranzo et al. (2008) and Abowd et al. (2003) use esti-

mates of person effects to measure the human capital distribution within firms. Combes et al. (2008) use a similar decomposition to estimate the contribution of neighborhoods to spatial earnings dispersion. Schmutte (2010) relies on consistent estimates of firm effects to infer the role of local job referral networks on earnings outcomes. Our estimates should be of interest in all such applications, as the specter of endogenous mobility clouds the interpretation of empirical results that rely on consistency of estimated person- or firm-effects.

There is also a parallel literature in the economics of education that uses value-added models to estimate the contributions of teachers and classrooms to student achievement. Endogenous assignment is just as much a problem there as it is here. Indeed, several recent studies have shown that the assumption of exogenous assignment in value-added models is rejected by the data (Rothstein 2010; Koedel and Betts 2010). Nevertheless, validation studies have shown that estimates from the value-added models are significantly correlated with independent assessments of teacher productivity. Our method would provide a direct method of assessing whether correcting the endogeneity bias in value-added models would substantially change their results. Our method can be implemented as long as one has a data set in which the realized network of connections between individuals and groups is sufficiently detailed to provide identifying variation.

We proceed by setting up the log earnings decomposition proposed by Abowd et al. (1999) so that we can clearly articulate the nature of the endogenous mobility problem. Then we report the results of two tests of the exogenous mobility assumption conducted by Abowd et al. (2010). As we will see, the null of exogenous mobility is rejected, but the associated analysis of mobility pat-



terns provides interesting information about the nature of the true model. Next, we present an illustrative theoretical model with endogenous job mobility and suggest an IV estimator based on the use of relational information. Turning to implementation, we set up a formal statistical model with endogenous mobility, in which earnings and job mobility are determined by latent classifications of workers, firms, and matches. This is a mixture model in which the probability of forming a link between a given worker and firm depends on a latent classification. We show that the model is identified and show how to estimate it using the Gibbs sampler.

## 4.2 What is Endogenous Mobility and Does it Matter?

Abowd et al. (1999) originally proposed the linear decomposition of log wage rates as the least squares fit of the equation

$$w = X\beta + D\theta + F\psi + \varepsilon \quad (4.1)$$

where  $w$  is the  $[N \times 1]$  stacked vector of log wage outcomes  $w_{it}$ ,  $X$  is the  $[N \times k]$  design matrix of observable individual and employer time-varying characteristics (the intercept is normally suppressed, with  $y$  and  $X$  measured as deviations from overall means);  $D$  is the  $[N \times I]$  design matrix for the individual effects;  $F$  is the  $[w \times J - 1]$  design matrix for the employer effects (non-employment is suppressed here).  $\varepsilon$  is the  $[N \times 1]$  vector of statistical errors whose properties will be elaborated below;  $\begin{bmatrix} \beta^T & \theta^T & \psi^T \end{bmatrix}^T$  are the unknown effects with dimension  $[k \times 1]$ ,  $[I \times 1]$ , and  $[J - 1 \times 1]$  associated with each of the design matrices.

The assumption of exogenous mobility appears in the restriction that  $E(\varepsilon|X, D, F) = 0$ . As long as the matrix of data moments has full rank – a non-

trivial assumption – this conditional moment restriction yields a consistent estimator for the parameter vector. But this imposes that a workers employment history is completely independent of the idiosyncratic part of earnings captured in  $\varepsilon$ . Intuitively, this model is equivalent to assuming that all assignments are pre-determined at birth given full knowledge of  $X, D, F$  and  $\begin{bmatrix} \beta^T & \theta^T & \psi^T \end{bmatrix}^T$ . Hence, there is no room for features included in most reasonable models of job mobility and assignment to affect either the duration of matches or the assignment of workers to particular employers.

Identification in the statistical model of  $\begin{bmatrix} \beta^T & \theta^T & \psi^T \end{bmatrix}$  also requires that  $[X D F]^T [X D F]$  be of full rank. Abowd et al. (2002) showed that this condition is equivalent to connectedness of the realized mobility network constructed by connecting all workers to every employer they are observed to match with during the sample. This network is a static bipartite graph on worker and employer nodes. As we will see, our identification strategy also has an interpretation in terms of the realized mobility network. We use information in the realized mobility network that predicts employer assignments but that we assume is conditionally independent of earnings residuals achieve identification.

The least squares solution for the parameters in equation are orthogonality of each design matrix with respect to the estimated residual implying that the estimated effects are also orthogonal to the estimated residual.

$$\hat{\beta}' X' \hat{\varepsilon} = 0, \hat{\theta}' D' \hat{\varepsilon} = 0 \text{ and } \hat{\psi}' F' \hat{\varepsilon} = 0 \quad (4.2)$$

### 4.2.1 Tests of Endogenous Mobility

Abowd et al. (2010) develop two formal tests for endogenous job mobility and apply them using employer-employee matched data from the Longitudinal Employer-Household Dynamics (LEHD) program of the U.S. Census Bureau. Here we survey the basic nature of the tests and their results. Their tests exploit the implicit restriction that future assignments of workers to firms are uninformative about current earnings residuals. Under the null hypothesis of endogenous mobility, these future assignments have no predictive power with respect to the residual. The first test, "Test 1", checks whether a worker's future employers are independent of the current period residual. The second test, "Test 2", checks whether the future employees of a particular employer are predictive of the residuals on their current period wage payments. For a thorough description of the test statistics, see Appendix B.3

Both tests reject the null of endogenous mobility. The test statistic for Test 1 is  $X^2_{8,991} = 7,438,692$  with  $\Pr\{X^2_{\nu_2}\} < 0.001$ . The test statistic for Test 2 is  $X^2_{900} = 172,295$  with  $\Pr\{X^2_{\nu_2}\} < 0.001$ . These are consistent with the related tests in Rothstein (2010) that falsify the endogenous mobility assumption in value-added models for longitudinally linked education data.

Some figures associated with the tests show an interesting relationship between estimates firm effects and residuals. Figure 4.1 shows the probability of coming from a job in each decile of the firm-effect ( $\psi$  distribution conditional on the decile of the current period residual. These plots show that workers who earn very high or very low residuals on their next job are more likely to have come from low wage employers than workers who exit to jobs that pay an average residual. Workers who are in average residual jobs are more likely to have

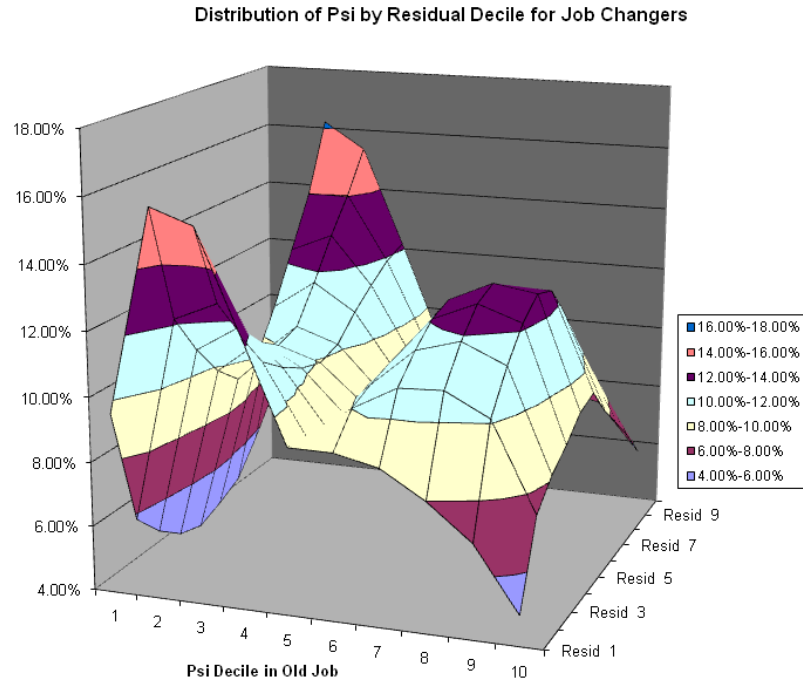


Figure 4.1: Test 1 of the exogenous mobility assumption: relationship between current period residual and past employer for workers who separate.

come from high wage jobs. Also, turning to figures 4.2-4.4 show frequencies of transition to employers in different deciles of the firm-effect distribution conditional on the current residual and  $\psi$ -decile. We see that the transition probabilities for workers with the lowest and highest residuals are far flatter than workers at the middle of the residual distribution.

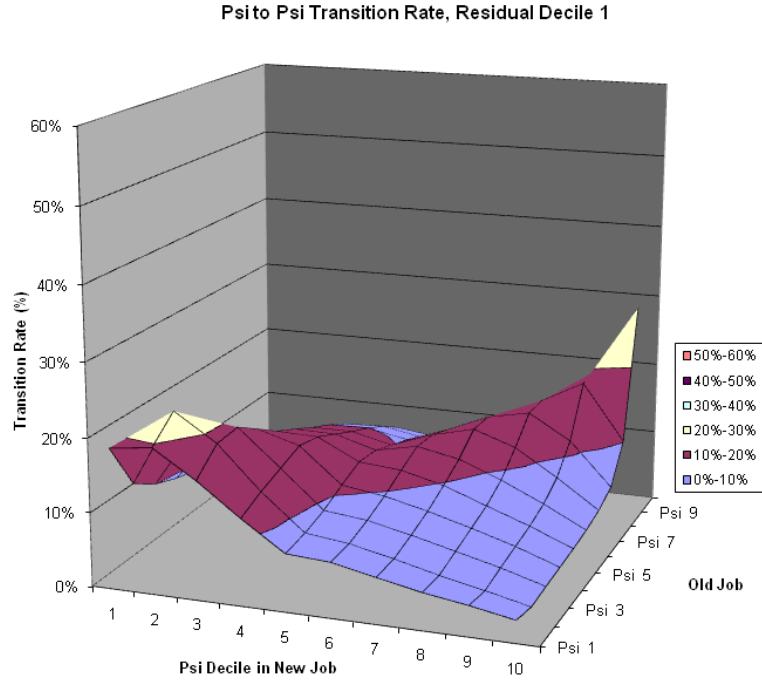


Figure 4.2: Test 1 of the exogenous mobility assumption: transition rates for workers with low earnings residuals

### 4.3 A Motivating Example of Endogenous Mobility with Comparative Advantage and Learning

To fix ideas regarding the endogeneity of job mobility and motivate our general estimation strategy, consider a labor market model with comparative advantage across firms and learning. The model is a variant of that proposed by Gibbons et al. (2005). The output of worker  $i$  working for employer  $j$  at time  $t$  is given by

$$y_{it} = \exp(X_{it}\beta + \varphi_{ijt}) \quad (4.3)$$

with  $\varphi_{ijt} = Z_i + b_j(\eta_i + \varepsilon_{ijt}) + c_j$

The components of the error term  $\varphi_{ijt}$  have familiar interpretations. The ef-

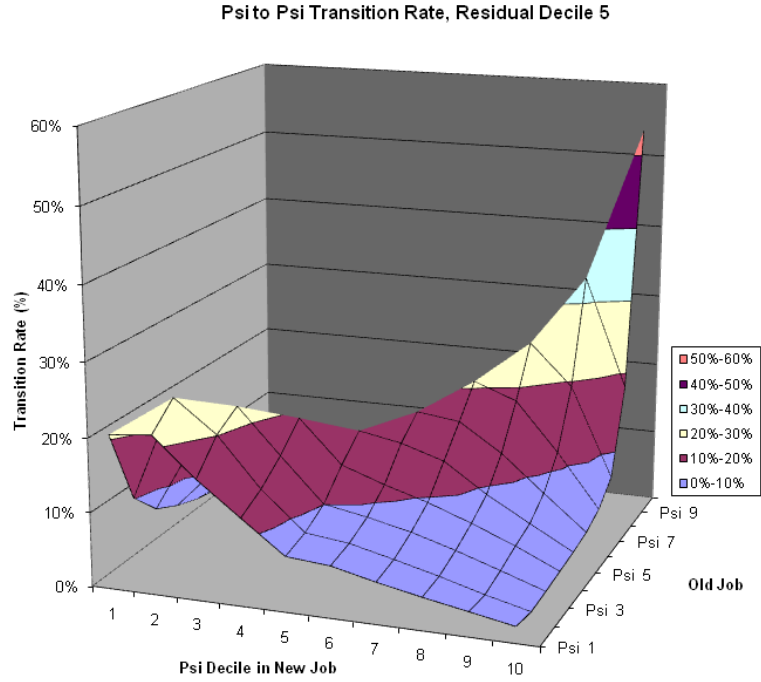


Figure 4.3: Test 1 of the exogenous mobility assumption: transition rates for workers with earnings residuals in the fifth decile

fect  $Z_i$  is the component of worker ability equally valued by all employers.  $\eta_i$  is an ability component that is differentially valued and assumed to be unknown to market participants.  $c_j$  is the true employer effect on productivity (and hence earnings) while  $b_j$  is an employer-specific valuation of  $\eta_i$ .

Assume that competition among employers is severe enough that workers receive a wage equal to their expected output. Under standard learning results, if output is commonly observed, then all market participants can update their beliefs about  $\eta_i$  from output realizations. Under mild restrictions on the relationship between  $b_j$  and  $c_j$ , workers will move from one employer to the next as beliefs about their ability evolve. In this setting, assignment to an employer with a particular realization of  $(c_j, b_j)$  is endogenous to the wage. In particu-

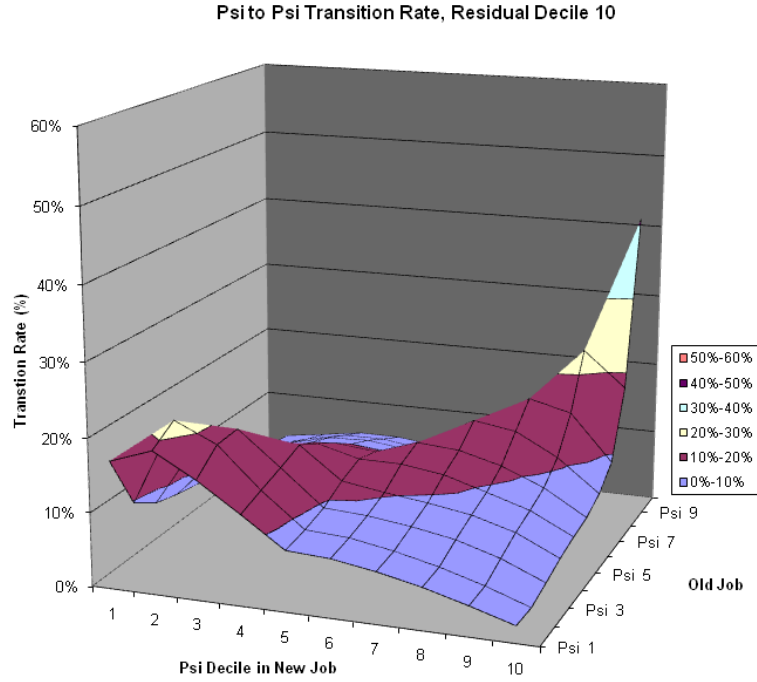


Figure 4.4: Test 1 of the exogenous mobility assumption: transition rates for workers with earnings residuals in the tenth decile

lar, positive output and wage shocks associated with employer learning will be correlated with mobility to employers with higher realization of  $b_j$ .

To pin things down a little more, note that when  $\varepsilon_{ijt}$  is log normal, we have:

$$\ln w_{it} = X_{it}\beta + Z_i + b_j m_{i,t-1} + c_j + \frac{1}{2} b_j^2 \sigma_t^2 + \mu_{it} \quad (4.4)$$

where  $m_{i,t-1}$  is the expectation after  $t-1$  output realizations of the unobserved component of worker ability.  $\sigma_t^2$  is the posterior variance of unobserved ability after  $t-1$  realizations. Finally,  $\mu_{it}$  is measurement error, which we assume to be orthogonal to sector affiliation along with everything else in the model. Details of the model may be found in Gibbons et al. (2005).

This provides a compelling case for why the Abowd et al. (1999) (hereafter

AKM) decomposition of log earnings into components associated with person and firm specific heterogeneity might be problematic in the presence of endogenous mobility. AKM estimate

$$\ln w_{it} = X_{it}\beta + \theta_i + \psi_{J(i,t)} + v_{it} \quad (4.5)$$

In this model, the error term embeds all of the latent heterogeneity associated with the learning process and its attenuation by employer-specific valuation of unknown ability components.

$$v_{it} = b_j m_{i,t-1} + \frac{1}{2} b_j^2 \sigma_i^2 + \mu_{it} \quad (4.6)$$

Under the Bayesian learning model, beliefs evolve as a Martingale so that we can express beliefs about worker  $i$ 's ability after  $t - 1$  periods,  $m_{i,t-1}$ , as the sum of initial beliefs and the sum of independent innovations:

$$m_{i,t-1} = m_{i,0} + \sum_{\tau=1}^{t-1} \xi_{i,\tau} \quad (4.7)$$

So we have

$$v_{it} = b_j m_{i,0} + b_j \sum_{\tau=1}^{t-1} \xi_{i,\tau} + \frac{1}{2} b_j^2 \sigma_i^2 + \mu_{it} \quad (4.8)$$

For the estimates of  $\theta$  and  $\psi$  from the AKM decomposition to be consistent for fixed ability,  $Z_i$ , and the employer specific productivity effect  $c_j$ , would require that employer assignment be orthogonal to  $v_{it}$ . Clearly, this is violated, since higher realizations of  $\xi_{i\tau}$  will be associated with assignment to employers with higher realizations of  $b_j$ . To the extent that  $c_j$  is negatively correlated with  $b_j$ , this will probably induce an attenuating bias on the true employer effect. In an extreme case, one could even imagine that this would invert the ordinality of the estimated employer effects  $\psi$  from the true ordering by values of  $c_j$ .



We can obtain consistent estimates provided an instrument is available for the endogenous assignments of workers to firms that is orthogonal to  $v_{it}$  (we also need a proxy for  $\frac{1}{2}b_j^2\sigma_i^2$ , but assume following Gibbons, et al. (2005) that this is wiped out by the inclusion of tenure and experience in  $X_{it}$ ). We can assume that initial beliefs  $m_{i,0}$  have mean zero conditional on observables since anything else can be absorbed into  $Z_i$ . The major problem is to find an instrument orthogonal to  $\sum_{\tau=1}^{t-1} \xi_{i,\tau}$ . Gibbons, et al. use lagged assignments and their interaction with observable characteristics as an instrument for endogenous sectoral assignments in a quasi-differenced version of the wage equation. Their procedure works because their method of quasi-differencing removes all of the lagged innovations  $\xi_{i\tau}$  except for one so that assignments two periods back are plausibly exogenous with respect to recent innovations. To maintain the AKM decomposition, we need an instrument that is orthogonal to the full history of innovations for each worker.

A natural candidate for such an instrument under the model is the set of assignments of your past coworkers, particularly those with similar observable characteristics. The model predicts that the market will hold similar beliefs about their unobserved productivities to your own, so in expectation, their future assignments should be good predictors of your own assignment, at least in terms of predicting the value of  $(b_j c)$ . At the same time, the model assumes that output realizations are independent across workers, so your innovations have nothing to do with your former coworkers' current employer assignments.

Intuitively this instrument will work because the innovations driving assignment will load error correlated with  $\eta_i$  onto the employers of your coworkers. Since their assignments are orthogonal to the innovations, they will infuse vari-

ance but not bias onto their  $\psi$  estimates. On average, most of these employers will be of a similar type to your own, so the employer-specific wage components,  $c_j$  will tend to get assigned to the 'right' types of employers.

Let  $\tilde{F}_{it}$  be the instrument vector with dimension identical to  $F_{it}$  which is the  $it$ 'th row of the design matrix of employer effects in AKM. The entries in  $\tilde{F}$  are the empirical probabilities of being employed by employer  $j$  in period  $t$  formed as the fraction of former coworkers with similar observable characteristics working in each of those firms. By construction,  $\tilde{F}$  orthogonal to  $v_{it}$  and correlated with  $F$ . The IV estimator solves

$$\begin{pmatrix} X'X & X'D & X'F \\ D'X & D'D & D'F \\ \tilde{F}'X & \tilde{F}'D & \tilde{F}'F \end{pmatrix} \begin{pmatrix} \beta \\ \theta \\ \psi \end{pmatrix} = \begin{pmatrix} X' \\ D' \\ \tilde{F}' \end{pmatrix} y \quad (4.9)$$

Where for the moment we ignore the potential endogeneity of participation and experience manifested in the design matrix of person effects,  $D$ . Implementation of the IV estimator turns on whether solving this system is computationally feasible.

#### 4.4 A General Model with Endogenous Mobility

In appendix B.2, we sketch the construction of an instrumental variables estimator based on the intuition derived in the previous section. That estimator takes advantage of the full structure of the realized mobility network. The realized mobility network has node sets representing firms and workers, with edge formation determined by a mobility process that depends on edge characteristics, including match quality, and node characteristics, including latent ability

of workers and latent productivity of employers.

Here, we set up a somewhat simplified version of that strategy. We assume workers, firms and matches have latent classifications that we cannot observe, but that determine earnings and mobility. The statistical model is very general and is compatible with many different structural models. Formally, we use the latent class model to identify workers and firms with similar mobility and earnings patterns and then to estimate the effect on earnings related to membership in these classes. We conduct Bayesian inference using an adaptation of the Gibbs sampler algorithm for finite mixture models (Tanner 1996; Diebolt and Robert 1994) to our case with multiple overlapping levels of correlation across observations. Our application and proposed procedure are related to stochastic blockmodels and other methods for the detection of ‘communities’ of nodes in social networks, the main innovation being our ability to use both node and edge characteristics in predicting the matches (Hoff et al. 2002; Newman and Leicht 2007; Neville and Jensen 2005).

#### 4.4.1 Model Setup

Agents of the model are workers, indexed  $i \in \{1 \dots I\} \equiv I$  and firms, indexed  $j \in \{0 \dots J\} \equiv J$ , where  $j = 0$  is “not employed.” Each worker has a latent ability class, denoted  $a_i \in A$ , and each firm, except  $j = 0$ , has a latent productivity class denoted  $b_j \in B$ . In addition, each worker-firm match has an associated heterogeneity component that affects both wages and mobility:  $k_{ij} \in K$ .  $A, B$  and  $K$  are discrete with cardinality  $L, M + 1$  and  $Q$ . The “not employed firm” is a single entity in its heterogeneity class, so the class  $b_0$  has no employer hetero-

geneity. To make the subsequent formulas easier to interpret, assume that the elements of  $A, B$  and  $K$  are rows from the identity matrices  $I_L, I_M$  and  $I_Q$ , respectively. For instance, if  $L = 2$ , we have  $A = \{(1, 0), (0, 1)\}$ . The assignments of workers and firms to ability and productivity classes are independent multinomial random variables with parameters  $\pi_a, \pi_b$ . We allow for endogeneity in the match quality by letting the probability of  $k$  depend on ability and productivity. So  $\Pr(k_{ij} = k | a_i = a, b_j = b) = \pi_{k|ab}$ .

The log of earnings, when actually employed in any match, is given by

$$w_{ijt} = \alpha + a_i\theta + b_j\psi + k_{ij}\mu + \varepsilon_{it} \quad (4.10)$$

where  $\theta, \psi, \mu$  are vectors of parameters describing the effect on the level of log earnings associated with membership in the various heterogeneity classes. We take  $\varepsilon$  to be normal with mean 0 and variance  $\sigma^2$ , independent and identically distributed across individuals and over time. When not employed, the individual earns a reservation log wage of

$$w_{i0t} = \alpha + a_i\theta + \psi_0 + k_{i0}\mu + \varepsilon_{it} \quad (4.11)$$

where  $\psi_0$  is just the appropriate element of  $\psi$  and  $k_{i0}\mu$  allows for heterogeneity in home production with the same effects as in the market sector.

We formalize endogenous mobility by allowing those matches and employment durations that are observed to depend on ability, productivity and match quality. Let  $J(i, t)$  be the index function that returns the identifier of the firm in which  $i$  is employed in period  $t$ . Define the variable  $s_{it} = 1$  if  $i$  separates from his current job at the end of period  $t$  and  $s_{it} = 0$  otherwise. We let the probability of separation depend on the match quality by specifying

$$\Pr(s_{it} = 1 | k_{iJ(i,t)}) = f_{se}(a_i, b_{J(i,t)}, k_{iJ(i,t)}; \gamma) \equiv \gamma_{abk} \quad (4.12)$$

where  $0 \leq \gamma_{abk} \leq 1$ . Conditional on separation, the productivity class of the next employer depends on the productivity of the current employer, the ability of the worker, and the quality of the current match

$$\Pr(b_{J(i,t+1)}|a_i, b_{J(i,t)}, k_{iJ(i,t)}) = f_{tr}(a_i, b_{J(i,t)}, k_{iJ(i,t)}; \delta) \equiv \delta_{abk} \in \Delta^{M+1} \quad (4.13)$$

where  $\delta_{abk} \equiv [\delta_{0|abk}, \dots, \delta_{M|abk}]$  is a  $1 \times (M+1)$  vector of transition probabilities,  $\Delta^{M+1}$  is the unit simplex, and  $J(i, 0) = 0$  for all  $i$ . The transition probabilities are indexed by all of the latent heterogeneity in the model. Within a heterogeneity class, the identity of the precise employer selected is completely random, as is the identity of an individual within an ability class.

## 4.5 Likelihood, Prior and Posterior Distributions

### 4.5.1 Likelihood functions

We begin by developing the likelihood function for the observed and latent data. The observed data,  $y_{it}$ , consist of wage rates, separations, accessions, and identifier information:

$$y_{it} = [w_{iJ(i,t)t}, s_{it}, i, J(i, t)] \text{ for } i = 1, \dots, I \text{ and } t = 1, \dots, T. \quad (4.14)$$

The latent data vector,  $Z$ , consists of the heterogeneity classifications.

$$Z = [a_1, \dots, a_I, b_1, \dots, b_J, k_{11}, k_{12}, \dots, k_{1J}, k_{21}, \dots, k_{IJ}] \quad (4.15)$$

In practice, we only ever use or update the heterogeneity classifications for the matches that are actually observed, the number of which is bounded above by  $T \times I$ . That is, we only care about  $k_{ij}$  where  $ij$  is such that  $j = J(i, t)$  for some  $t$ .

Finally, the complete parameter vector is

$$\rho^T = [\alpha, \theta^T, \psi^T, \mu^T, \sigma, \gamma, \delta, \pi_a, \pi_b, \pi_{k|ab}], \rho \in \Theta \quad (4.16)$$

We assume that workers and firms are infinitely-lived. The complete process starts at  $t = 1$ , with continuous sampling continuing to date  $T$ . We model initial conditions by assuming that everyone enters the labor force at  $t = 1$  and are assigned an employer completely at random. In other words, we assume that the matches initially observed are exogenous. The observed data matrix for this time interval is denoted  $Y$ . The likelihood function for the joint distribution  $(Y, Z)$  is given by

$$\mathfrak{L}(\rho|Y, Z) \propto \prod_{i=1}^I \left\{ \prod_{t=1}^T \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[ -\frac{(w_{iJ(i,t)} - \alpha - a_i\theta - b_{J(i,t)}\psi - k_{iJ(i,t)}\mu)^2}{2\sigma^2} \right] \right. \\ \left. \times \prod_{t=1}^{T-1} \left[ 1 - \gamma_{\langle a_i \rangle \langle b_{J(i,t)} \rangle \langle k_{iJ(i,t)} \rangle} \right]^{1-s_{it}} \left[ \gamma_{\langle a_i \rangle \langle b_{J(i,t)} \rangle \langle k_{iJ(i,t)} \rangle} \right]^{s_{it}} \right. \\ \left. \times \prod_{t=1}^{T-1} \left[ \delta_{\langle b_{J(i,t+1)} \rangle | \langle a_i \rangle \langle b_{J(i,t)} \rangle \langle k_{iJ(i,t)} \rangle} \right]^{s_{it}} \right\} \quad (4.17)$$

$$\times \prod_{i=1}^I \prod_{j=1}^J \left[ \left( \prod_{\ell=1}^L (\pi_{a\ell})^{a_{i\ell}} \right) \left( \prod_{m=1}^M (\pi_{bm})^{b_{jm}} \right) \left( \prod_{q=1}^Q (\pi_{q|\ell m})^{k_{ijq}} \right) \right] \quad (4.18)$$

where the notation  $\pi_{a\ell}$  denotes the  $\ell^{th}$  element of  $\pi_a$  (similarly for  $\pi_{bm}$ ,  $b_{jm}$ , etc.) and  $\langle x \rangle$  means the index of the non-zero element of the vector  $x$ .

The likelihood factors into a part due to the observed data conditioned on the latent, and the latent conditioned on the parameters.

The observed data likelihood conditional on the latent data factors further into separate contribution from earnings and the mobility process. The mobility process is Markov, and conditionally independent of the earnings realizations once we know the latent classifications of the workers firms and matches.

The power of the model comes from the predictive equation for  $Z$  given the observed data and the parameters, which we can compute as the complete data

likelihood divided by the observed data likelihood. The observed data likelihood is calculated by margining out the latent data.

### 4.5.2 Prior distributions

The parameter vector  $\rho$  has a prior distribution that is composed of the product of priors on each of the main components of the parameter space. Conditional on the heterogeneity probabilities, the coefficients in the log wage equation have prior distributions proportional to a constant (each one uniform on a wide, but finite, interval of  $\mathbb{R}$ ) and subject to the constraint that the probability-weighted average effects are all zero. That is,

$$\pi_a^T \theta = \pi_b^T \psi = \pi_{k|ab(\ell m)}^T \mu = 0 \quad (4.19)$$

for all  $\ell, m$  where  $\pi_{k|ab(\ell m)} \equiv \Pr(k_{ij} = k | a_i = \ell, b_j = m)$ . The variance parameter,  $\sigma$ , has the inverted gamma prior  $\text{IG}(\nu_0, s_0)$  with prior degrees of freedom small and prior  $s_0^2$  large. Each vector of probabilities has a Dirichlet prior with each element of the parameter vector given by the inverse of the dimension of the probability vector.

### 4.5.3 Posterior distributions

The posterior distribution of  $\rho$  given  $(Y, Z)$  is given by

$$\begin{aligned} p(\rho | Y, Z) &\propto \mathcal{L}(\rho | Y, Z) \frac{1}{\sigma^{\nu_0+1}} \exp\left(-\frac{s_0^2}{\sigma^2}\right) \prod_{\ell=1}^L \pi_{a\ell}^{\frac{1}{L}-1} \prod_{m=1}^M \pi_{bm}^{\frac{1}{M}-1} \\ &\times \prod_{\ell=1}^L \prod_{m=0}^M \prod_{q=1}^Q \left( \pi_{q|\ell m}^{\frac{1}{Q}-1} \gamma_{\ell m q}^{\frac{1}{2}-1} (1 - \gamma_{\ell m q})^{\frac{1}{2}-1} \prod_{m'=0}^M \delta_{m'|\ell m q}^{\frac{1}{M+1}-1} \right). \end{aligned} \quad (4.20)$$

Which factors into independent posterior distributions as follows:

$$\begin{bmatrix} \alpha \\ \theta \\ \psi \\ \mu \end{bmatrix} | \sigma \sim N \left( \begin{bmatrix} \hat{\alpha} \\ \hat{\theta} \\ \hat{\psi} \\ \hat{\mu} \end{bmatrix}, \sigma^2 (G^T G)^{-1} \right) \quad (4.21)$$

where

$$\begin{bmatrix} \hat{\alpha} \\ \hat{\theta} \\ \hat{\psi} \\ \hat{\mu} \end{bmatrix} = (G^T G)^{-1} G^T w$$

$$\sigma^2 \sim \text{IG} \left( \frac{\nu}{2}, \frac{2}{\nu s^2} \right) \quad (4.22)$$

$$\pi_a \sim \text{D} \left( n_{a1} + \left( \frac{1}{L} - 1 \right), \dots, n_{aL} + \left( \frac{1}{L} - 1 \right) \right) \quad (4.23)$$

$$\pi_b \sim \text{D} \left( n_{b1} + \left( \frac{1}{M} - 1 \right), \dots, n_{bM} + \left( \frac{1}{M} - 1 \right) \right) \quad (4.24)$$

$$\pi_{k|ab} \sim \text{D} \left( n_{k|ab1} + \left( \frac{1}{Q} - 1 \right), \dots, n_{k|abQ} + \left( \frac{1}{Q} - 1 \right) \right) \quad (4.25)$$

$$\gamma_{lmq} \sim \text{D} \left( n_{\ell mq}^{sep} + \left( \frac{1}{2} - 1 \right), n_{\ell mq}^{stay} + \left( \frac{1}{2} - 1 \right) \right) \quad (4.26)$$

$$\delta_{b|lmq} \sim \text{D} \left( n_{0|\ell mq}^{trans} + \left( \frac{1}{M+1} - 1 \right), \dots, n_{M|\ell mq}^{trans} + \left( \frac{1}{M+1} - 1 \right) \right) \quad (4.27)$$

The above factorizations contain some new notation.  $G = [A \ B \ K]$  is the full design matrix of ability, productivity, and match types in the complete data,



and  $w$  is the vector of observed earnings. The term  $v$  in the posterior of  $\sigma$  is  $v = N + v_0 - (L + M + Q)$  and

$$s^2 = \frac{\left( w - G \begin{bmatrix} \hat{\alpha} \\ \theta \\ \psi \\ \mu \end{bmatrix} \right)^T \left( w - G \begin{bmatrix} \hat{\alpha} \\ \theta \\ \psi \\ \mu \end{bmatrix} \right)}{v} \quad (4.28)$$

the rest of the parameters are sampled from Dirichlet posteriors, denoted by  $D$ . Finally, we have various counts from the completed data.  $n_{a\ell}$  is the count of workers with ability class  $\ell$ .  $n_{bm}$  is the number of employers in productivity class  $m$ .  $n_{k|abq}$  is the number of matches observed in quality class  $q$ .  $n_{lmq}^{sep}$  is the number of observations in which a worker in ability class  $\ell$  separates from an employer in productivity class  $m$  when match quality was  $q$ . Finally,  $n_{m'\ell mq}^{trans}$  is the number of transitions by workers in ability class  $\ell$  from a match with an employer in productivity class  $m$  and match quality class  $q$  to an employer in productivity class  $m'$ .

## 4.6 Estimation Procedure

We start with initial values for the parameter vector and latent data,  $\rho^{(0)}, Z^{(0)}$ . We have already defined the distributions of the parameters given the observed and latent data above. To complete the specification, we define the distributions for the latent variables conditional on the observed data and the parameters. For instance, to update the ability classifications for the workers, we need to sample

from a multinomial with probability of the  $\ell^{th}$  class equal to

$$\begin{aligned} p(a_i = \ell | a_{-i}, b, k, Y, \rho) &= \frac{p(a_{-i}, b, k, Y | \rho, a_i = \ell) p(a_i = \ell)}{p(a_{-i}, b, k, Y | \rho)} \\ &= \frac{\pi_{a\ell} p(a_{-i}, b, k, Y | \rho, a_i = \ell) p(a_i = \ell)}{\sum_{\ell'=1}^L [\pi_{a\ell'} p(a_{-i}, b, k, Y | \rho, a_i = \ell') p(a_i = \ell')]} \end{aligned} \quad (4.29)$$

This requires computing the likelihood function under each assignment of  $i$  to an ability classification. The update formulas for  $b_j$  and  $k_{ij}$  are exactly analogous. This is a high dimension procedure, requiring roughly  $L$  evaluations of the likelihood per individual,  $M$  per firm, and  $Q$  per match, for each iteration. However, given the simple form of the likelihood, these computations should not be excessively burdensome. Furthermore, much of this work can be parallelized if necessary. The updating for each  $a_i$  is an independent task. Furthermore, most of the structure of the likelihood function remains the same as we tweak individual assignments, which we exploit to obtain further simplification.

With the posterior distributions as defined in the previous section, the Gibbs sampler can be implemented as follows:

$$\begin{aligned} \sigma^{(1)} &\sim p(\sigma | \alpha^{(0)}, \theta^{(0)T}, \psi^{(0)T}, \mu^{(0)T}, \gamma^{(0)}, \delta^{(0)}, \pi_a^{(0)}, \pi_b^{(0)}, \pi_{k|ab}^{(0)}, a_1^{(0)}, \dots, k_{IJ}^{(0)}, Y) \\ \begin{bmatrix} \alpha \\ \theta \\ \psi \\ \mu \end{bmatrix}^{(1)} &\sim p \left( \begin{bmatrix} \alpha \\ \theta \\ \psi \\ \mu \end{bmatrix} \middle| \gamma^{(0)}, \delta^{(0)}, \pi_a^{(0)}, \pi_b^{(0)}, \pi_{k|ab}^{(0)}, a_1^{(0)}, \dots, k_{IJ}^{(0)}, \sigma^{(1)}, Y \right) \\ \gamma^{(1)} &\sim p(\gamma | \delta^{(0)}, \pi_a^{(0)}, \pi_b^{(0)}, \pi_{k|ab}^{(0)}, a_1^{(0)}, \dots, k_{IJ}^{(0)}, \alpha^{(1)}, \theta^{(1)T}, \psi^{(1)T}, \mu^{(1)T}, Y) \\ &\vdots \\ a_I^{(1)} &\sim p(a_I | \delta^{(1)}, \pi_a^{(1)}, \pi_b^{(1)}, \pi_{k|ab}^{(1)}, a_1^{(1)}, \dots, a_{I-1}^{(1)}, b_1^{(1)}, \dots, k_{IJ}^{(1)}, \alpha^{(1)}, \theta^{(1)T}, \psi^{(1)T}, \mu^{(1)T}, Y) \end{aligned}$$

## 4.7 Monte Carlo Estimates

We demonstrate the validity of our estimation procedure with a Monte Carlo study. Our model is an extension of standard data augmentation (Tanner 1996). As such, theoretical results on the convergence of the data augmentation algorithm and Gibbs sampler should obtain here. The main novelty in our approach is the presence of multiple levels of latent variables. Each observation belongs to three separate latent classes: a worker-specific ability class, an employer-specific productivity class, and a match-specific quality class. The model thus poses problems of both identification and of convergence. As we show, our Gibbs sampler performs well for inference regarding the parameters associated with earnings effects of the latent classifications. Specifically, our model detects endogenous mobility bias that is the object of primary interest in this study.

To demonstrate our estimation procedure, we simulate data under a model with  $L = M = Q = 2$  heterogeneity classes. In our model economy, there are  $I = 100$  workers and  $J = 20$  employers. We observe workers in each of  $T = 50$  time periods. The simulated data allow for both the separation decision and the job allocation at transition to depend on match quality. Furthermore, match quality is correlated with latent worker ability and latent employer productivity, so there is a rich structure of endogenous mobility in the model. We simulate the model under the parameterization described in section B.1.1.

Section B.1.2 shows summary statistics for the data simulated under the model. The data contain 373 distinct employer-employee matches, not including spells of unemployment. The simulated job mobility and earnings histories reflect the model parameterization. As expected given the model, the average

of log earnings per period, calculated across periods of employment, is indistinguishable from zero.

Of primary interest for our purpose are estimates of the Abowd et al. (1999) decomposition in these data. We use the preconditioned conjugate gradient algorithm to solve the fixed effect moment equations with person- and firm-effects, as in Abowd et al. (2002). The correlation between estimated person- and firm-effects on realized matches in these estimates is 0.12. The true correlation based on the actual values of  $\theta$  and  $\psi$  used to generate the data is 0.10.

We implement the Gibbs sampler to estimate this model. To implement the sampler, we need to choose starting points for the latent heterogeneity classes as well as for the model parameters. Furthermore, we need to choose a burn-in period to allow the sampler to converge to the posterior. In this Monte Carlo exercise, the sampler converges within 100 iterations to point mass on the true latent classifications. Thereafter, the sampler converges immediately to sampling from the posterior distribution for the parameters. This is because the posterior for the parameters given the latent data factors into the product of Dirichlet and Normal-Gamma posteriors. The results presented here use randomly chosen starting values for the latent heterogeneity classes. We initialize all parameters in the wage equation to zero, and start the categorical variables at uniform distributions. The sampler iterates between sampling from the posterior distribution for the model parameters conditional on the observed and latent data, and sampling from the predictive distribution for the latent classes given the observed data and parameters. The choice of starting point does not matter: up to a relabeling of the heterogeneity classes, the sampler quickly converges to the same solution.

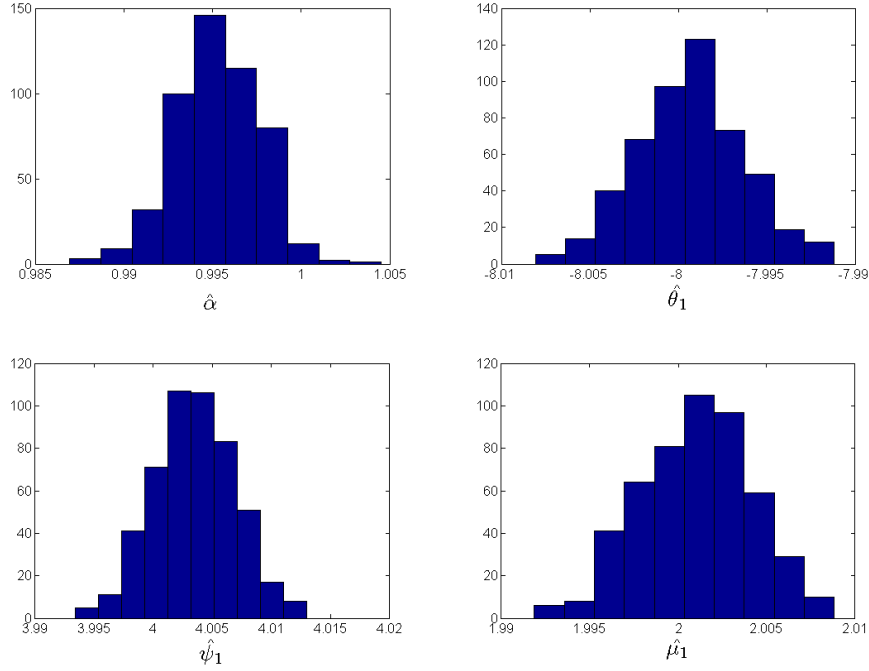


Figure 4.5: Posterior distribution of parameters from the earnings equation based on draws from the Gibbs sampler.

To sample from the posterior for  $\alpha, \theta, \psi, \mu, \sigma$  requires an identifying assumption. Here, unlike Abowd et al. (2002), we set one effect equal to zero. Since we only have two classes for each type of heterogeneity, this means we estimate four parameters for the earnings equation along with the standard deviation of the structural error. In addition, we have to specify prior parameters for the Normal-Gamma distribution. We assume prior degrees of freedom equal to one, and also set the prior standard error equal to one. The prior for  $(\alpha, \theta, \psi, \mu)$  is normal with mean zero and prior covariance matrix is the identity matrix.

Figure 4.5 contains histogram plots of the posterior of the parameters from the wage equation. By comparing these values with the truth presented in section B.1.2, we see that up to reidentification, the model converges to the truth.

As a result, our procedure is able to detect the small endogeneity bias in the estimated correlation between worker and employer heterogeneity classes. Finally, section B.1.3 presents the posterior mode for the categorical variables in the model. A comparison with the model parameters and the simulated separation, transition, and match rates show that the model is very accurately capturing the evidence from the data. Again, this is unsurprising since the simulator quickly finds the true latent ability, productivity, and match classifications for the data.

## 4.8 Conclusion

Could we randomly assign workers to jobs without changing the manner in which their wages were determined? Either their employers know of the random assignment, and would, presumably, compensate them differently than workers they hired, or they would not, in which case those workers would be non-randomly selected from the pool of potential applicants. It is easy to imagine randomizing applications but not realized assignments. We do not have an ideal experiment that identifies the effect of assignments of workers to firms. It is difficult to think of what an ideal experiment would be.

The central problem of this paper is one manifestation of a fundamental challenge of empirical social science: separating the influence of correlated unobservables and sorting from the direct effect of group membership. Exploiting the wealth of information about labor market behavior locked in the relational structure of matched data holds great potential to address these problems. We use matched data to construct an instrument for the actual assignment of workers to firms that exploits the relational structure of our data. The key insight

is that the work histories of one's coworkers and previous employers are informative of one's own employment history, while being plausibly unrelated to whatever idiosyncratic wage innovations drive assignment at the margin.

APPENDIX A

JOB REFERRAL NETWORKS AND THE DETERMINATION OF  
EARNINGS IN LOCAL LABOR MARKETS

### A.1 Proof of Proposition 1

**Proof.** These results are standard in the job search literature. See Rogerson, Shimer and Wright (2005). Since workers are wealth maximizers, and the evolution of portable skills  $e_{it}$  is unrelated to  $p_{J(i,t)}$ , we can model search over wage premia,  $p$ , and ignore determination of  $e$ . Since workers are myopic about the evolution of the referral network, the decision environment is stationary so the value of holding a job with wage premium  $p$  is given by the Bellman equation

$$rV(p) = p + \lambda_1 \int_0^\infty [\max\{V(p'), V(p)\} - V(p)] d\tilde{F}(p') + \delta [U - V(p)] \quad (\text{A.1})$$

where  $r$  is the discount rate,  $U$  is the value of becoming unemployed, and  $\tilde{F}(p)$  is the cumulative distribution of offers,  $p$ , appropriately transformed from  $F(\psi)$  given above. The key behavioral assumption is that workers act as if the offer distribution  $\tilde{F}$  is fixed. The corresponding Bellman equation for the value of becoming unemployed is

$$rU = p_b + \lambda_0 \int_0^\infty [\max\{V(p') - U\}] d\tilde{F}(p') \quad (\text{A.2})$$

It is clear that  $V(p)$  is increasing in  $p$  and that  $U$  is constant. Therefore, employed workers will adopt a strategy where they exit unemployment whenever  $p > p_R$  for some constant  $p_R$  and switch jobs whenever they receive an offer with  $p' > p$ . The reservation premium,  $p_R$  will satisfy

$$p_R = p_b + (\lambda_0 - \lambda_1) \int_{p_R}^\infty \left[ \frac{1 - \tilde{F}(p)}{r + \delta + \lambda_1 [1 - \tilde{F}(p)]} dp \right] \quad (\text{A.3})$$



■

## A.2 Proof of proposition 3

**Proof.** For the proof, I suppress dependence on  $Z$ ,  $W$  and  $\Psi$ . It is also convenient to introduce notation. As suggested in the statement of the proposition, stars added to a distribution indicate that they are the truncated versions of the unstarred distribution. For instance,  $g^*(\psi) = g(\psi|\psi > \psi_0)$  Some algebra reveals that the truncated mean is a mixture

$$\begin{aligned} E_{f^*}(\psi) &= a^* E_{g^*}(\psi) + (1 - a^*) E_{h^*}(\psi) \\ &= a^* \mu_{g^*} + (1 - a^*) \mu_{h^*} \end{aligned}$$

where  $a^* = \frac{a(1-G(\psi_0))}{1-aG(\psi_0)-(1-a)H(\psi_0)}$  Taking derivatives

$$\frac{\partial \mu_{f^*}}{\partial \mu_h} = \frac{\partial a^*}{\partial \mu_h} \mu_{g^*} + a^* \frac{\partial \mu_{g^*}}{\partial \mu_h} - \frac{\partial a^*}{\partial \mu_h} \mu_{h^*} + (1 - a^*) \frac{\partial \mu_{h^*}}{\partial \mu_h}$$

Eliminating  $\frac{\partial \mu_{g^*}}{\partial \mu_h}$  and rearranging

$$\frac{\partial \mu_{f^*}}{\partial \mu_h} = \frac{\partial a^*}{\partial \mu_h} (\mu_{g^*} - \mu_{h^*}) + (1 - a^*) \frac{\partial \mu_{h^*}}{\partial \mu_h}$$

Log concavity of  $h$  ensures  $\frac{\partial \mu_{h^*}}{\partial \mu_h} > 0$ . Furthermore it is clear that  $\frac{\partial a^*}{\partial \mu_h} > 0$ . Thus, as long as

$$\left| (\mu_{g^*} - \mu_{h^*}) \right| < \frac{(1 - a^*) \frac{\partial \mu_{h^*}}{\partial \mu_h}}{\frac{\partial a^*}{\partial \mu_h}}$$

we have  $\frac{\partial \mu_{f^*}}{\partial \mu_h} > 0$  ■

### A.3 Proof of proposition 4

**Proof.** The  $q$ th quantile of the distribution of observed offers,  $\psi^q$  is defined implicitly by

$$\int^{\psi^q} f(\psi|\psi > \psi_0) = \int_{\psi_0}^{\psi^q} \frac{f(\psi)}{1 - F(\psi_0)} d\psi = q$$

Which gives

$$F(\psi^q) = q + (1 - q)F(\psi_0)$$

Renormalize the offer distribution in terms of deviations from its mean,  $\mu$ :

$$F(\psi^q - \mu) = q + (1 - q)F(\psi_0 - \mu)$$

First, consider the effect of a shift in the initial offer on the  $q$ th quantile of observed jobs

$$F'(\psi^q - \mu) \frac{\partial \psi^q}{\partial \psi_0} = (1 - q)F'(\psi_0 - \mu).$$

This establishes that a shift in the initial offer is expected to have a positive effect on all quantiles of the observed offer distribution. The goal is to assess how the magnitude of this effect varies with respect to the quantile  $q$ . Hence, we want to establish the sign of

$$\frac{\partial^2 \psi^q}{\partial q \partial \psi_0}$$

Note

$$\frac{\partial \psi^q}{\partial q} = \frac{1 - F(\psi_0 - \mu)}{F'(\psi^q - \mu)}$$

differentiating this with respect to  $\psi_0$

$$\frac{\partial^2 \psi^q}{\partial q \partial \psi_0} = \frac{-F'(\tilde{\psi}_0) - F''(\tilde{\psi}^q) \frac{\partial \psi^q}{\partial q} \frac{\partial \psi^q}{\partial \psi_0}}{F'(\tilde{\psi}^q)}$$

where I have replaced  $\psi - \mu = \tilde{\psi}$  for simplicity

$$= -\frac{F'(\tilde{\psi}_0)}{F'(\tilde{\psi}^q)} - \frac{F''(\tilde{\psi}^q)(1 - q)(1 - F(\tilde{\psi}_0))}{F'(\tilde{\psi}^q)^3}.$$

When  $F''(\tilde{\psi}^q) > 0$ , this is negative. Suppose  $F''(\tilde{\psi}^q) < 0$ . I will show that  $\frac{\partial^2 \psi^q}{\partial q \partial \psi_0} > 0$  is impossible as long as

$$\frac{F'(\tilde{\psi}^q)^2}{|F''(\tilde{\psi}^q)|} \geq 1 - F(\tilde{\psi}^q)$$

This condition simply places limits on the amount of curvature in the density function. Note that in the case described in the statement of the proposition, where  $F'$  is a symmetric density function and  $F$  is log concave, we have

$$\frac{F'(\psi^q - \mu)^2}{|F''(\psi^q - \mu)|} = \frac{F'(\mu - \psi^q)^2}{F''(\mu - \psi^q)} \geq F(\mu - \psi^q) = 1 - F(\psi^q - \mu)$$

where the first and last equalities follow by symmetry of the density function, the inequality follows from log concavity. For details on log concave functions and their application to search models, see Bergstrom and Bagnoli, 2005 or Flinn and Heckman 1983.

Continuing with the proof, suppose  $F''(\tilde{\psi}^q) < 0$  and  $\frac{\partial^2 \psi^q}{\partial q \partial \psi_0} > 0$ . Then

$$\frac{-F''(\tilde{\psi}^q)(1 - q)(1 - F(\tilde{\psi}_0))}{F'(\tilde{\psi}^q)^2} > 1$$

that is

$$\frac{F'(\tilde{\psi}^q)^2}{|F''(\tilde{\psi}^q)|} < (1 - q)(1 - F(\tilde{\psi}_0))$$

which by the assumption above implies

$$\begin{aligned} 1 - F(\tilde{\psi}^q) &< (1 - q)(1 - F(\tilde{\psi}_0)) \\ 1 - (q + (1 - q)F(\tilde{\psi}_0)) &< (1 - q)(1 - F(\tilde{\psi}_0)) \\ (1 - q)(1 - F(\tilde{\psi}_0)) &< (1 - q)(1 - F(\tilde{\psi}_0)), \end{aligned}$$

a contradiction. It follows that  $\frac{\partial^2 \psi^q}{\partial q \partial \psi_0} < 0$ .

The proof that  $\frac{\partial^2 \psi^q}{\partial q \partial \mu} > 0$  is analogous. ■

APPENDIX B  
ENDOGENOUS MOBILITY (WITH JOHN M. ABOWD)

**B.1 Details of the Monte Carlo Estimates**

**B.1.1 Parameters for the Simulated Data**

This section presents the parameters used to generate simulated data. The notation is as in the main body of the paper. We use the notation 0 as the employer productivity class label during spells of unemployment. So, the notation  $\delta_{10}$  denotes the vector of destination probabilities for a worker with ability type 1 who was unemployed. Note also that the columns of  $\delta$  are ordered so that the probability of transition to employers of productive type 1 and 2 appear in the first and second columns. The third column is the probability of transition to unemployment.

$$\pi_A = (0.50, 0.50)$$

$$\pi_B = (0.50, 0.50)$$

$$\pi_{K|AB} = \begin{bmatrix} \pi_{K|11} \\ \pi_{K|12} \\ \pi_{K|21} \\ \pi_{K|22} \end{bmatrix} = \begin{bmatrix} 0.50, 0.50 \\ 0.25, 0.75 \\ 0.75, 0.25 \\ 0.50, 0.50 \end{bmatrix}$$

$$\gamma = \begin{bmatrix} \gamma_{111} \\ \gamma_{112} \\ \gamma_{121} \\ \gamma_{122} \\ \gamma_{10-} \\ \gamma_{211} \\ \gamma_{212} \\ \gamma_{221} \\ \gamma_{222} \\ \gamma_{20-} \end{bmatrix} = \begin{bmatrix} 0.250 \\ 0.175 \\ 0.100 \\ 0.025 \\ 0.600 \\ 0.150 \\ 0.075 \\ 0.025 \\ 0.025 \\ 0.500 \end{bmatrix}$$

$$\delta = \begin{bmatrix} \delta_{111} \\ \delta_{112} \\ \delta_{121} \\ \delta_{122} \\ \delta_{10-} \\ \delta_{211} \\ \delta_{212} \\ \delta_{221} \\ \delta_{222} \\ \delta_{20-} \end{bmatrix} = \begin{bmatrix} b=1 & b=2 & b=U \\ 0.4 & 0.1 & 0.5 \\ 0.4 & 0.4 & 0.2 \\ 0.6 & 0.2 & 0.2 \\ 0.3 & 0.5 & 0.2 \\ 0.5 & 0.5 & 0.0 \\ 0.5 & 0.2 & 0.3 \\ 0.2 & 0.7 & 0.1 \\ 0.5 & 0.4 & 0.1 \\ 0.1 & 0.8 & 0.1 \\ 0.3 & 0.7 & 0.0 \end{bmatrix}$$

$$\theta = (4, -4)$$

$$\psi = (2, -2)$$

$$\mu = (1, -1)$$

$$\alpha = 0$$

$$\sigma = 0.1$$

## B.1.2 Summary of the Simulated Data

- Match Rates:

$$\begin{bmatrix} m_{K|11} \\ m_{K|12} \\ m_{K|21} \\ m_{K|22} \end{bmatrix} = \begin{bmatrix} 0.55, 0.45 \\ 0.24, 0.76 \\ 0.77, 0.23 \\ 0.54, 0.46 \end{bmatrix}$$

- Separation rates:

$$\begin{bmatrix} s_{111} \\ s_{112} \\ s_{121} \\ s_{122} \\ s_{10-} \\ s_{211} \\ s_{212} \\ s_{221} \\ s_{222} \\ s_{20-} \end{bmatrix} = \begin{bmatrix} 0.270 \\ 0.145 \\ 0.090 \\ 0.026 \\ 0.621 \\ 0.143 \\ 0.092 \\ 0.023 \\ 0.033 \\ 0.578 \end{bmatrix}$$

- Transition Rates:

$$\begin{bmatrix} d_{111} \\ d_{112} \\ d_{121} \\ d_{122} \\ d_{10-} \\ d_{211} \\ d_{212} \\ d_{221} \\ d_{222} \\ d_{20-} \end{bmatrix} = \begin{array}{c|cc} & b = 1 & b = 2 & b = U \\ \hline & 0.30 & 0.10 & 0.60 \\ & 0.35 & 0.51 & 0.14 \\ & 0.83 & 0.00 & 0.17 \\ & 0.31 & 0.44 & 0.25 \\ & 0.46 & 0.54 & 0.00 \\ & 0.49 & 0.20 & 0.31 \\ & 0.06 & 0.89 & 0.06 \\ & 0.70 & 0.22 & 0.87 \\ & 0.09 & 0.82 & 0.09 \\ & 0.31 & 0.69 & 0.00 \end{array}$$

### B.1.3 Posterior Mode

$$\hat{\pi}_A = (0.55, 0.45)$$

$$\hat{\pi}_B = (0.50, 0.50)$$

$$\hat{\pi}_{K|AB} = \begin{bmatrix} \hat{\pi}_{K|11} \\ \hat{\pi}_{K|12} \\ \hat{\pi}_{K|21} \\ \hat{\pi}_{K|22} \end{bmatrix} = \begin{bmatrix} 0.55, 0.46 \\ 0.25, 0.75 \\ 0.76, 0.24 \\ 0.56, 0.44 \end{bmatrix}$$

$$\hat{\gamma} = \begin{bmatrix} \hat{\gamma}_{111} \\ \hat{\gamma}_{112} \\ \hat{\gamma}_{121} \\ \hat{\gamma}_{122} \\ \hat{\gamma}_{10-} \\ \hat{\gamma}_{211} \\ \hat{\gamma}_{212} \\ \hat{\gamma}_{221} \\ \hat{\gamma}_{222} \\ \hat{\gamma}_{20-} \end{bmatrix} = \begin{bmatrix} 0.27 \\ 0.14 \\ 0.09 \\ 0.03 \\ 0.62 \\ 0.14 \\ 0.09 \\ 0.02 \\ 0.03 \\ 0.58 \end{bmatrix}$$



$$\hat{\delta} = \begin{bmatrix} \hat{\delta}_{111} \\ \hat{\delta}_{112} \\ \hat{\delta}_{121} \\ \hat{\delta}_{122} \\ \hat{\delta}_{10-} \\ \hat{\delta}_{211} \\ \hat{\delta}_{212} \\ \hat{\delta}_{221} \\ \hat{\delta}_{222} \\ \hat{\delta}_{20-} \end{bmatrix} = \begin{bmatrix} b=1 & b=2 & b=U \\ 0.30 & 0.10 & 0.60 \\ 0.35 & 0.51 & 0.14 \\ 0.83 & 0.00 & 0.17 \\ 0.31 & 0.44 & 0.25 \\ 0.46 & 0.54 & 0.00 \\ 0.49 & 0.20 & 0.31 \\ 0.06 & 0.89 & 0.06 \\ 0.70 & 0.22 & 0.08 \\ 0.09 & 0.82 & 0.09 \\ 0.31 & 0.69 & 0.00 \end{bmatrix}$$

$$\hat{\theta} = (-8, 0)$$

$$\hat{\psi} = (4, 0)$$

$$\hat{\mu} = (2, 0)$$

$$\hat{\alpha} = 1$$

$$\hat{\sigma} = 0.1$$

## B.2 A Proposed IV Estimator Based on Exploitation of the Full Realized Mobility Network

### B.2.1 The IV Estimator

- Organize the data to allow application of the Abowd et al. (2002) solver for the LS fixed-effects design

- Solve

$$\begin{bmatrix} X'X & X'D & X'F \\ D'X & D'D & D'F \\ F'X & F'D & F'F \end{bmatrix} \begin{bmatrix} \beta \\ \theta \\ \psi \end{bmatrix}_{(0)} = \begin{bmatrix} X'y \\ D'y \\ F'y \end{bmatrix} \quad (\text{B.1})$$

- Store  $\begin{bmatrix} \hat{\beta}' & \hat{\theta}' & \hat{\psi}' \end{bmatrix}_{(0)}$

### First Reduced Form

- Discretize  $\theta_{(0)}$  onto  $L$  fixed points of support. Discretize  $\psi_{(0)}$  onto  $M$  fixed points of support.
- Fit a reduced form random graph from adjacency matrices of the discrete distributions of  $\theta$  and  $\psi$ .
- The reduced form random graph model estimates the transition matrix

$$\Pr \left[ b_{rs}(t) = 1 | b_{ij}(t-1) = 1 \text{ for } i, r = 1, \dots, L; j, s = 1, \dots, M \right] \quad (\text{B.2})$$

- Details of the estimation of the transition matrix follow the algorithm
- Simulate  $\tilde{B}(t)$  and  $\tilde{D}(t)$   $R \times (L + M)$  times. For  $t = 1, \dots, T$  form

$$\hat{B}(t) = \frac{1}{R} \sum_{r=1}^R \tilde{B}_r(t) \quad \hat{D}(t) = \frac{1}{R} \sum_{r=1}^R \tilde{D}_r(t) \quad (\text{B.3})$$

- Define the instrument matrices

$$\hat{D} = \begin{bmatrix} \hat{D}(1) \\ \hat{D}(2) \\ \vdots \\ \hat{D}(T) \end{bmatrix} \quad \hat{F} = \begin{bmatrix} \hat{B}(1) \\ \hat{B}(2) \\ \vdots \\ \hat{B}(T) \end{bmatrix} \quad (\text{B.4})$$

### First IV Step

- Solve

$$\begin{bmatrix} \beta \\ \theta \\ \psi \end{bmatrix}_{(1)} = \arg \min \left\{ \begin{array}{c} (y - X\beta - D\theta - F\psi)' \begin{bmatrix} X & \hat{D} & \hat{F} \end{bmatrix} \\ \begin{bmatrix} X'X & X'\hat{D} & X'\hat{F} \\ \hat{D}'X & \hat{D}'\hat{D} & \hat{D}'\hat{F} \\ \hat{F}'X & \hat{F}'\hat{D} & \hat{F}'\hat{F} \end{bmatrix}^{-1} \\ \begin{bmatrix} X & \hat{D} & \hat{F} \end{bmatrix}' (y - X\beta - D\theta - F\psi) \end{array} \right\} \quad (\text{B.5})$$

- Notice that the instrument matrix

$$\begin{bmatrix} X & \hat{D} & \hat{F} \end{bmatrix} \begin{bmatrix} X'X & X'\hat{D} & X'\hat{F} \\ \hat{D}'X & \hat{D}'\hat{D} & \hat{D}'\hat{F} \\ \hat{F}'X & \hat{F}'\hat{D} & \hat{F}'\hat{F} \end{bmatrix}^{-1} \begin{bmatrix} X & \hat{D} & \hat{F} \end{bmatrix}' \quad (\text{B.6})$$

has column rank at most  $k + L + M$  whereas the original LS fixed-effects problem had column rank at most  $k + I + J - 1$ .

- The IV estimator for the individual and employer heterogeneity only identifies effects with dimensionality given by the fixed discretization.
- The IV estimates are consistent in the presence of endogenous mobility but can be improved by refitting the reduced form random graph model.
- Store  $\begin{bmatrix} \hat{\beta}' & \hat{\theta}' & \hat{\psi}' \end{bmatrix}_{(1)}$ .

### Second Reduced Form

- Discretize  $\hat{\theta}_{(1)}$  onto  $L$  fixed points of support. Discretize  $\hat{\psi}_{(1)}$  onto  $M$  fixed points of support. Use the same support points as in step 0.
- Re-fit the reduced form random graph using the discrete distributions of  $\theta$  and  $\psi$ .

## Second IV Step

- Re-simulate  $\tilde{B}(t)$  and  $\tilde{D}(t)$   $R \times (L + M)$  times. For  $t = 1, \dots, T$  form

$$\hat{B}(t) = \frac{1}{R} \sum_{r=1}^R \tilde{B}_r(t) \quad \hat{D}(t) = \frac{1}{R} \sum_{r=1}^R \tilde{D}_r(t) \quad (\text{B.7})$$

from the newly simulated adjacency matrices.

- Define new instrument matrices

$$\hat{D} = \begin{bmatrix} \hat{D}(1) \\ \hat{D}(2) \\ \vdots \\ \hat{D}(T) \end{bmatrix} \quad \hat{F} = \begin{bmatrix} \hat{B}(1) \\ \hat{B}(2) \\ \vdots \\ \hat{B}(T) \end{bmatrix} \quad (\text{B.8})$$

- Solve

$$\begin{bmatrix} \beta \\ \theta \\ \psi \end{bmatrix}_{(2)} = \arg \min \left\{ \begin{array}{c} (y - X\beta - D\theta - F\psi)' \begin{bmatrix} X & \hat{D} & \hat{F} \end{bmatrix} \\ \begin{bmatrix} X'X & X'\hat{D} & X'\hat{F} \\ \hat{D}'X & \hat{D}'\hat{D} & \hat{D}'\hat{F} \\ \hat{F}'X & \hat{F}'\hat{D} & \hat{F}'\hat{F} \end{bmatrix}^{-1} \\ \begin{bmatrix} X & \hat{D} & \hat{F} \end{bmatrix}' (y - X\beta - D\theta - F\psi) \end{array} \right\} \quad (\text{B.9})$$

- The resulting  $\hat{\theta}_{(2)}$  and  $\hat{\psi}_{(2)}$  are consistent in the presence of endogenous mobility and efficient if the bipartite random graph model is true.

$$p(\{\mathbf{B}\}_{t=1}^T | \Theta) \equiv \text{pdf of bipartite adjacency matrix} \quad (\text{B.10})$$

$$p(\Theta) \equiv \text{pdf of parameter prior} \quad (\text{B.11})$$

So the Markov chain is defined by

$$p\left(\tilde{\Theta} \left| \left\{ \tilde{\mathbf{B}} \right\}_{t=1}^T, \{\ln \mathbf{w}_i\}_{t=1}^T, \mathbf{X} \right.\right) \quad (\text{B.12})$$

$$p\left(\left\{ \tilde{\mathbf{B}} \right\}_{t=1}^T, \{\ln \mathbf{w}_i\}_{t=1}^T \left| \tilde{\Theta}, \mathbf{X} \right.\right) \quad (\text{B.13})$$

Initial estimation of  $\Theta$  from  $\{\mathbf{B}\}_{t=1}^T$  observed. Then, put workers and firms into classes (initially 100 each). In the simulation, the problem is to assign workers to new firms without locking them into the original (initial) classification. Proposed strategy is to do a preliminary inconsistent estimation of AKM decomposition, then fill the current 100 category classification, run the simulator, draw samples of  $\{\tilde{\mathbf{B}}\}_{t=1}^T$  at convergence. Then for each sampled  $\{\tilde{\mathbf{B}}\}_{t=1}^T$ , convert back to the original identifier set by a random model with probabilities developed as follows: put the people back where they came from (same  $i$ ), use the estimated mobility model to determine if there has been a transition, if not leave  $j$  unchanged. Otherwise, sample  $j$  from the correct current group. Refit the AKM model with  $\tilde{Z}$  averaged over all the draws from the current iteration.

### B.3 Formulas for Endogeneity Tests

These formulas are derived and discussed thoroughly in Abowd et al. (2010).

#### B.3.1 Test 1

$$\bar{\varepsilon}_{it-1} = \frac{\sum_{\{s|J(i,s)=j \wedge s < t \wedge J(i,s) \neq J(i,t)\}} \widehat{\varepsilon}_{is}}{\sum 1 \{s|J(i,s) = j \wedge s < t \wedge J(i,s) \neq J(i,t)\}} \quad (\text{B.14})$$

(average residual for the most recent completed job at  $j$  by  $i$ )

For all  $(i, t)$  where  $J(i, t-1) \neq J(i, t)$  (job changers) compute the counts

$$n_{abcd} = \sum_{\{i,t|J(i,t-1) \neq J(i,t)\}} 1 \left\{ Q(\theta_i) = a \wedge Q(\psi_{J(i,t-1)}) = b \wedge Q(\psi_{J(i,t)}) = c \wedge Q(\bar{\varepsilon}_{it-1}) = d \right\} \quad (\text{B.15})$$

and let

$$\pi_{abcd} = \Pr \left\{ Q(\theta_i) = a \wedge Q(\psi_{J(i,t-1)}) = b \wedge Q(\psi_{J(i,t)}) = c \wedge Q(\bar{\varepsilon}_{it-1}) = d \right\} \quad (\text{B.16})$$

then

$$X_{\nu_1}^2 = \text{Test}(\pi_{abcd} = \pi_{abc+} \pi_{++d}) \quad (\text{B.17})$$

where  $\nu_1 = (\#(Q(\theta_i)) \times \#Q(\psi_{J(i,t-1)}) \times \#Q(\psi_{J(i,t)}) - 1) \times (\#Q(\bar{\varepsilon}_{it-1}) - 1)$

### B.3.2 Test 2

$$\widetilde{\varepsilon}_{jt} = \frac{\sum_{\{i|J(i,t)=j\}} \widehat{\varepsilon}_{it}}{\sum 1_{\{i|J(i,t)=j\}}} \quad (\text{B.18})$$

(average residual for all employees at  $j$  in year  $t$ )

For two periods  $s < t$  and all firms alive in period  $s$ , compute the counts

$$n_{abc|s} = \sum_j \left\{ 1 \left\{ Q(\psi_j) = a \wedge Q(\widetilde{\varepsilon}_{js}) = c \right\} \times \sum_{\{i|J(i,s)=j\}} Q(\theta_i) = b \right\} \quad (\text{B.19})$$

and

$$n_{abc|t} = \sum_j \left\{ 1 \left\{ Q(\psi_j) = a \wedge Q(\widetilde{\varepsilon}_{jt}) = c \right\} \times \sum_{\{i|J(i,t)=j\}} Q(\theta_i) = b \right\}. \quad (\text{B.20})$$

Note that the two counts are not independent because they condition on the same distribution of employers alive in period  $s$ . Let:

$$\pi_{abc|s} = \Pr \left\{ Q(\psi_j) = a \wedge (Q(\theta_i) = b|s) \wedge Q(\tilde{\varepsilon}_{js}) = c \right\} \quad (\text{B.21})$$

and

$$\pi_{abc|t} = \Pr \left\{ Q(\psi_j) = a \wedge (Q(\theta_i) = b|t) \wedge Q(\tilde{\varepsilon}_{js}) = c \right\} \quad (\text{B.22})$$

then

$$X_{\nu_2}^2 = Test \left( \ln \left( \frac{\pi_{abc|s}}{\pi_{abc|t}} \right) = \ln \left( \frac{\pi_{ab+|s}}{\pi_{ab+|t}} \right) \right) \quad (\text{B.23})$$

where  $\nu_2 = (\#Q(\theta_i) - 1) \times (\#Q(\tilde{\varepsilon}_{js}) - 1) + (\#Q(\psi_j) - 1) \times (\#Q(\theta_i) - 1) \times (\#Q(\tilde{\varepsilon}_{js}) - 1)$ .

Proof

$$\begin{aligned} \ln \left( \frac{\pi_{abc|s}}{\pi_{abc|t}} \right) &= (\mu_{as} - \mu_{at}) + (\mu_{bs} - \mu_{bt}) + (\mu_{cs} - \mu_{ct}) \\ &\quad + (\gamma_{abs} - \gamma_{abt}) + (\gamma_{acs} - \gamma_{act}) + (\gamma_{bcs} - \gamma_{bct}) \\ &\quad + (\rho_{abcs} - \rho_{abct}) \end{aligned}$$

where

$$(\mu_{as} - \mu_{at}) = \text{change in main effects of } Q(\psi_j) \equiv 0$$

$$(\mu_{bs} - \mu_{bt}) = \text{change in main effects of } Q(\theta_i), \text{ with } df = (\#Q(\theta_i) - 1)$$

$$(\mu_{cs} - \mu_{ct}) = \text{change in main effects of } Q(\tilde{\varepsilon}_{js}) \equiv 0$$

$$(\gamma_{abs} - \gamma_{abt}) = \text{change in interaction } Q(\psi_j) \text{ and } Q(\theta_i), \text{ with } df = (\#Q(\psi_j) - 1) \times (\#Q(\theta_i) - 1)$$

$$(\gamma_{acs} - \gamma_{act}) = \text{change in interaction } Q(\psi_j) \text{ and } Q(\tilde{\varepsilon}_{js}) \equiv 0$$

$$(\gamma_{bcs} - \gamma_{bct}) = \text{change in interaction } Q(\theta_i) \text{ and } Q(\tilde{\varepsilon}_{js}), \text{ with } df = (\#Q(\theta_i) - 1) \times (\#Q(\tilde{\varepsilon}_{js}) - 1)$$

$$(\rho_{abcs} - \rho_{abct}) = \text{change in interaction of } Q(\psi_j), Q(\theta_i) \text{ and } Q(\tilde{\varepsilon}_{js}), \text{ with } df = (\#Q(\psi_j) - 1) \times (\#Q(\theta_i) - 1) \times (\#Q(\tilde{\varepsilon}_{js}) - 1)$$

Under the null hypothesis  $(\gamma_{bcs} - \gamma_{bct}) = 0$  and  $(\rho_{abcs} - \rho_{abct}) = 0$ , the change in  $df$  is  $(\#Q(\theta_i) - 1) \times (\#Q(\tilde{\varepsilon}_{js}) - 1) + (\#Q(\psi_j) - 1) \times (\#Q(\theta_i) - 1) \times (\#Q(\tilde{\varepsilon}_{js}) - 1)$ .



## BIBLIOGRAPHY

- Abowd, J. M., Creecy, R. H. and Kramarz, F. (2002). Computing person and firm effects using linked longitudinal employer-employee data, *Technical Report TP-2002-06*, LEHD, U.S. Census Bureau.
- Abowd, J. M., Kramarz, F. and Margolis, D. N. (1999). High wage workers and high wage firms, *Econometrica* **67**(2): 251–333.
- Abowd, J. M., Kramarz, F., Pérez-Duarte, S. and Schmutte, I. M. (2009). A formal test of assortative matching in the labor market. NBER Working Paper No. 15546.
- Abowd, J. M., Lengermann, P. and McKinney, K. L. (2003). The measurement of human capital in the U.S. economy, *Technical Report TP-2002-09*, LEHD U.S. Census Bureau.
- Abowd, J. M., McKinney, K. and Schmutte, I. M. (2010). How important is endogenous mobility for measuring employer and employee heterogeneity? mimeo.
- Abraham, K. G. and Farber, H. S. (1987). Job duration, seniority, and earnings, *The American Economic Review* **77**(3): 278–297.
- Bayer, P., Ross, S. and Topa, G. (2008). Place of work and place of residence: Informal hiring networks and labor market outcomes, *Journal of Political Economy* **116**: 1150–1194.
- Bertrand, M., Luttmer, E. F. P. and Mullainathan, S. (2000). Network effects and welfare culture, *Quarterly Journal of Economics* **115**: 1019–1055. Priority: normal.

- Bewley, T. F. (1999). *Why Wages Don't Fall During a Recession*, Harvard University Press.
- Bjelland, M., Fallick, B., Haltiwanger, J. and McInterfer, E. (2008). Employer-to-employer flows in the united states: estimates using linked employer-employee data, *Technical Report Working Paper 13867*, National Bureau of Economic Research.
- Blume, L. E. and Durlauf, S. N. (2005). Identifying social interactions: A review. July 22, 2005.
- Bramoullé, Y., Djebbari, H. and Fortin, B. (2009). Identification of peer effects through social networks, *Journal of Econometrics* **150**: 41–55.
- Brock, W. A. and Durlauf, S. N. (2001). Interactions-based models, *Handbook of Econometrics*, Elsevier.
- Buchinsky, M., Fougère, D., Kramarz, F. and Tchernis, R. (2003). Interfirm mobility, wages, and the returns to seniority in the U.S. Working Paper.
- Burdett, K. and Mortensen, D. T. (1998). Wage differentials, employer size, and unemployment, *International Economic Review* **39**: 257–273.
- Cahuc, P. and Fontaine, F. (2002). On the efficiency of job search with social networks, *Technical Report 583*, IZA.
- Cahuc, P., Postel-Vinay, F. and Robin, J.-M. (2006). Wage bargaining with on-the-job search: Theory and evidence, *Econometrica* **74**(2): 323–364.
- Calvo-Armengol, A. and Jackson, M. O. (2004). The Effects of Social Networks on Employment and Inequality, *American Economic Review* **94**(3).

- Calvo-Armengol, A. and Jackson, M. O. (2007). Networks in labor markets: Wage and employment dynamics and inequality, *Journal of Economic Theory* **132**(1): 27–46.
- Cameron, C. A., Gelbach, J. B. and Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors, *Review of Economics and Statistics* **90**: 414–427.
- Case, A. E. and Katz, L. F. (1991). The company you keep: The effects of family and neighborhood on disadvantaged youths, *Technical report*, National Bureau of Economic Research.
- Clauset, A., Newman, M. and Moore, C. (2004). Finding community structure in very large networks, *Physical Review E* **70**: 066111.
- Combes, P.-P., Duranton, G. and Gobillon, L. (2008). Spatial wage disparities: Sorting matters!, *Journal of Urban Economics* **63**: 723–742.
- Conley, T. G. and Topa, G. (2002a). Socio-economic distance and spatial patterns in unemployment, *Journal of Applied Econometrics* **17**: 303–327.
- Conley, T. G. and Topa, G. (2002b). Socio-economic distance and spatial patterns in unemployment, *Journal of Applied Econometrics* **17**: 303–327.
- Conley, T. G. and Topa, G. (2007). Estimating dynamic local interactions models, *Journal of Econometrics* **140**: 282–303.
- Conley, T. G. and Udry, C. R. (2010). Learning about a new technology: Pineapple in Ghana, *American Economic Review* **100**: 35–69.
- Copic, J., Jackson, M. O. and Kirman, A. (2009). Identifying community struc-

- tures from network data via maximum likelihood methods, *The B.E. Journal of Theoretical Economics* **9**(1): Article 30.
- Davidoff, T. (2005). Income sorting: Measurement and decomposition, *Journal of Urban Economics* **58**: 289–303.
- de Melo, R. L. (2008). Sorting in the labor market: Theory and measurement. Working Paper.
- Diebolt, J. and Robert, C. P. (1994). Estimation of finite mixture distributions through Bayesian sampling, *Journal of the Royal Statistical Society. Series B (Methodological)* **56**(2): 363–375.
- Fontaine, F. (2007). A job search model with social networks: The better match hypothesis. September 19, 2007.
- Gibbons, R., Katz, L. F., Lemieux, T. and Parent, D. (2005). Comparative advantage, learning and sectoral wage determination, *Journal of Labor Economics* **23**: 681–723.
- Girvan, M. and Newman, M. (2002). Community structure in social and biological networks, *Proceedings of the National Academy of Sciences of the United States of America* **99**: 7821–7826.
- Gruetter, M. and Lalive, R. (2009). The importance of firms in wage determination, *Labour Economics* **16**: 149–160.
- Guillaume, J.-L. and Latapy, M. (2004). Bipartite graphs as models of complex networks, *CAAN*, pp. 127–139.
- Heckman, J. J. and Honore, B. E. (1990). The empirical content of the Roy model, *Econometrica* **58**(5): 1121–1149.

- Hellerstein, J. K., McInerney, M. and Neumark, D. (2008). Measuring the importance of labor market networks, *Technical report*, National Bureau of Economic Research.
- Hoff, P. D., Raftery, A. and Handcock, M. (2002). Latent space approaches to social network analysis, *Journal of the American Statistical Association* **97**: 1090–1098.
- Holland, P. and Leinhardt, S. (1977). A dynamic model for social networks, *Journal of Mathematical Sociology* **5**: 5–20.
- Ioannides, Y. M. (2004). Neighborhood income distribution, *Journal of Urban Economics* **56**: 435–457.
- Ioannides, Y. M. and Loury, L. D. (2004). Job information networks, neighborhood effects, and inequality, *Journal of Economic Literature* **42**: 1056–1093. Priority: normal.
- Iranzo, S., Schivardi, F. and Tosetti, E. (2008). Skill dispersion and firm productivity: An analysis with matched employer-employee data, *Journal of Labor Economics* **26**: 247–285.
- Jackson, M. O. (2005). A survey of models of network formation: Stability and efficiency, <http://www.hss.caltech.edu/jacksonm/netsurv.pdf>.
- Kambourov, G. and Manovskii, I. (2008). Rising occupational and industrial mobility in the united states: 1968-1997, *International Economic Review* **49**(1): 41–79.
- Koedel, C. and Betts, J. R. (2010). Does student sorting invalidate value-added models of teacher effectiveness? an extended analysis of the Rothstein critique, *Education Finance and Policy* **5**: 54–81.

- Koskinen, J. H. and Snijders, T. A. (2007). Bayesina inference for dynamic social network data, *Journal of Statistical Planning and Inference* **137**(12): 3930–3938.
- Kremer, M. (1997). How much does sorting increase inequality, *Quarterly Journal of Economics* **112**: 115–39.
- Kremer, M. and Maskin, E. (1996). Wage inequality and segregation by skill, *Technical Report Working Paper 5718*, National Bureau of Economic Research.
- le Maire, D. and Scheur, C. (2008). Job sampling and sorting. Working Paper.
- Lentz, R. (Forthcoming). Sorting by search intensity, *Journal of Economic Theory* .
- Liben-Nowell, D. and Kleinberg, J. (2003). The link prediction problem for social networks, *Proc. 12th Int’l Conference on Information and Knowledge Management (CIKM)* .
- Lise, J., Meghir, C. and Robin, J.-M. (2008). Martching, sorting and wages. Working Paper.
- Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem, *Review of Economic Studies* **60**: 531–542.
- Montgomery, J. D. (1991). Social networks and labor market outcomes: Toward an economic analysis, *American Economic Review* **81**: 1408.
- Moreno, J. (1934). *Who Shall Survive? Foundations of Sociometry, Group Psychotherapy and Sociodrama*, Nervous and Mental Disease Publishing Company, Washington D.C., Washington, D.C.
- Mortensen, D. and Vishwanath, T. (1994). Personal contacts and earnings: It is who you know!, *Labour Economics* **1**: 187–201.

- Neville, J. and Jensen, D. (2005). Leveraging relational autocorrelation with latent group models, *MRDM'05, August 21, 2005, Chicago, Illinois, USA* pp. 49–55.
- Newman, M. (2003). The structure and function of complex networks, *SIAM Review* **45**(2): 167–256.
- Newman, M. E. J. and Leicht, E. A. (2007). Mixture models and exploratory analysis in networks, *Proceedings of the National Academy of Sciences of the United States* **104**: 9564–9569.
- Newman, M., Strogatz, S. H. and Watts, D. J. (2001). Random graphs with arbitrary degree distributions and their applications, *Physical Review E* **64**: 026118.
- Postel-Vinay, F. and Robin, J.-M. (2002). Wage dispersion with worker and employer heterogeneity, *Econometrica* **70**: 2295–350.
- Rogerson, R., Shimer, R. and Wright, R. (2005). Search-theoretic models of the labor market: A survey, *Journal of Economic Literature* **43**: 959–988.
- Rothstein, J. (2010). Teacher quality in educational production: Tracking, decay and student achievement, *Quarterly Journal of Economics* **125**(1): 175–214.
- Sarkar, P. and Moore, A. W. (2005). Dynamic social network analysis using latent space models, *SIGKDD Explorations* **7**(2): 31–40.
- Sattinger, M. (1993). Assignment models of the distribution of earnings, *Journal of Economic Literature* **XXXI**(2): 831–880.
- Schmutte, I. (2010). Job referral networks and the determination of earnings in local labor markets. Cornell University Department of Economics.

- Snijders, T. A. B. (2005). *Models for Longitudinal Network Data*, Cambridge University Press, New York.
- Soderberg, B. (2002). General formalism for inhomogeneous random graphs, *Physical Review E* **66**: 066121.
- Tanner, M. A. (1996). *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*, Springer Series in Statistics, 3rd edn, Springer-Verlag.
- Topa, G. (2001). Social interactions, local spillovers and unemployment, *Review of Economic Studies* **68**: 261–295.
- Wahba, J. and Zenou, Y. (2005). Density, social networks and job search methods: Theory and application to egypt, *Journal of Development Economics* **78**: 443–473.
- Wasserman, S. and Faust, K. (1994). *Social Network Analysis: Methods and Applications*, Cambridge University Press.
- Weinberg, B. A., Reagan, P. B. and Yankow, J. J. (2004). Do neighborhoods affect hours worked? evidence from longitudinal data, *Journal of Labor Economics* **22**: 891–924.
- Woodcock, S. (2008). Wage differentials in the presence of unobserved worker, firm, and match heterogeneity, *Labour Economics* **15**: 771–793.